"Who is he?" Event-related brain potentials and unbound pronouns

"Who is he?" Event-related brain potentials and unbound pronouns

"Who's he?" Event-related brain potentials and unbound pronouns

Mante S. Nieuwland

Department of Psychology, School of Philosophy, Psychology and Language Sciences,
University of Edinburgh, Edinburgh, United Kingdom

Keywords: Pronouns, Anaphora, Gender, Ambiguity, Sentence comprehension, ERPs, Nref, P600

Correspondence:
Mante S. Nieuwland
Department of Psychology
School of Philosophy, Psychology and Language Sciences
University of Edinburgh
7 George Square
Edinburgh EH8 9JZ
Scotland, United Kingdom
Email: m.nieuwland@ed.ac.uk
Telephone: +44 (0) 131 650 8387

ABSTRACT

Three experiments used event-related potentials to examine the processing consequences of gender-mismatching pronouns (e.g., "The aunt found out that he had won the lottery"), which have been shown to elicit P600 effects when judged as syntactically anomalous (Osterhout & Mobley, 1995). In each experiment, mismatching pronouns elicited a sustained, frontal negative shift (Nref) compared to matching pronouns: when participants were instructed to posit a new referent for mismatching pronouns (Experiment 1), and without this instruction (Experiment 2 and 3). In Experiment 1 and 2, the observed Nref was robust only in individuals with higher reading span scores. In Experiment 1, participants with lower reading span showed P600 effects instead, consistent with an attempt at coreferential interpretation despite gender mismatch. The results from the experiments combined suggest that, in absence of an acceptability judgment task, people are more likely to interpret mismatching pronouns as referring to an unknown, unheralded antecedent than as a grammatically anomalous anaphor for a given antecedent.

INTRODUCTION

Speakers tend to use pronouns (e.g., 'she' and 'it') to refer to entities that are in the focus of attention, sidestepping the costs associated with explicitly repeating antecedents (e.g, Almor & Nair, 2007; Ariel, 1988; Arnold, 2010; Gordon, Grosz & Gilliom, 1993; Gordon & Hendrick, 1998; Gundel, Hedberg & Zacharski, 1993). This preference is mirrored in comprehension: comprehenders have a strong preference for antecedents that are readily available (e.g., Garnham 2001; Garrod, O'Brien, Morris & Rayner, 1990; Gernsbacher, 1989; Sanford & Garrod, 1989). In the context of a readily available antecedent, unbound pronouns like 'he' and 'she' are often understood immediately, and without effort (e.g., Clark & Sengul, 1979), despite the fact that they are always formally ambiguous (Chomsky, 1981). In fact, an antecedent does not even have to be explicitly mentioned in order to be available for reference. Pronouns without explicit antecedents (i.e., unheralded; Gerrig, Horton & Stent 2011) are ubiquitous, usually when the intended referents are part of *common ground* (i.e., shared experience or sociocultural knowledge, providing background and antecedents for sentences such as "They dug up the road again"; Gerrig et al., 2011; Greene, Gerrig, MacKoon & Ratcliff, 1994; see also Kitzinger, Shaw & Toerien, 2012; Sanford, Garrod, Lucas & Henderson, 1983). This opens up the question, though, of how people understand pronouns for which the context does not provide an explicit, suitable antecedent. For example, in the sentence "The aunt found out that he had won the lottery," how is *he* interpreted? Who is *he* most likely to be?

Absence of a suitable antecedent could arise when speaker and listener are not talking about the same people or objects (see also Barr & Keysar, 2002), or when speakers make gender agreement errors (see Vigliocco & Franck, 1999). Furthermore, pronouns sometimes precede referents ('cataphoric pronouns', e.g., "While he was sleeping, John became very

rich"; see Gernsbacher & Jescheniak, 1995; Filik & Sanford, 2008; Kazanina, Lau, Lieberman, Yoshida & Phillips, 2007; Van Gompel & Liversedge, 2003). Resolution of a pronoun without a suitable antecedent may thus depend on whether one thinks that the speaker is talking about somebody new or that the speaker simply made a grammatical mistake. The current work examines the online processing consequences of such referentially problematic utterances, as reflected in the brain's electrophysiology. Whereas many electrophysiological studies on pronoun resolution have investigated the processing consequences of biological and/or grammatical gender mismatch (e.g., Hammer, Jansma, Lamer & Münte, 2008; Harris, Wexler & Holcomb, 2000; Lamers, Jansma, Hammer & Münte, 2006; Nieuwland & Van Berkum, 2006; Osterhout, Bersick & Mclaughlin, 1997a; Osterhout & Mobley, 1995; Streb, Hennighausen & Rösler, 2004; Qiu, Swaab, Chen & Wang, 2012; Xu, Jiang & Zhou, 2013), the fact that pronouns can introduce novel referents has been largely ignored in the study of pronoun resolution (but see Filik, Sanford & Leuthold, 2008).

Behavioral and ERP studies on pronoun resolution suggest that readers and listeners do not delay attempts to interpret a pronoun until it can be unambiguously resolved, and that resolution is rapidly shaped by pronominal gender (e.g., Arnold, Eisenband, Brown-Schmidt & Trueswell, 2000; Koornneef & Van Berkum, 2006; Kreiner, Sturt & Garrod, 2008; Nieuwland & Van Berkum, 2006; Osterhout et al., 1997a; Sturt, 2003). Crucially, this incrementality has potential implications for what happens when people read "The aunt found out that" followed by 'she' or 'he' (henceforth called *matching* and *mismatching* pronouns), even though both pronouns are a grammatical continuation. I will discuss two possible processing situations before mapping them onto extant theories of reference processing and spelling out predictions for brain responses.

The first possibility is that the pronoun - irrespective of its gender – in interpreted as referring to 'aunt'. This referentially parsimonious interpretation may be driven by the strong preference to have locally available, prominent referents (e.g., Foraker & McElree, 2007; see Cunnings, Patterson & Felser, 2014, for antecedent recency effects; for a review, see Garnham, 2001; Garrod & Sanford, 1994). This interpretation is not imposed by the syntactic constraints of English grammar (the pronoun is 'free' as opposed to reflexive pronouns that must be bound within a syntactically defined local domain; Chomsky, 1981). Nevertheless, if comprehenders try to establish a coreferential interpretation for 'he' and 'aunt', this might lead them to *perceive* a violation of the formal requirement that coreferring elements agree in gender. This possibility was advocated by Osterhout and Mobley (1995), who reported that such pronouns elicit ERPs that are also seen in response to outright syntactic violations (e.g., subject-verb number agreement). Arguably, when comprehenders perceive a syntactic anomaly, they can adhere to the initial coreferential interpretation by assuming that the wrong gender was used.

The second possibility is that people interpret 'she' as referring to 'aunt' but assume a novel, unknown referent for 'he'[1]. This may come about when pronoun gender is used as a non-defeasible filter on anaphoric interpretation (see Badecker & Straub 2002; Sturt, 2013). The invocation of a novel referent could be construed as an elaborative inference if readers go beyond what is literally stated and try to infer who the referent might plausibly be (e.g., Graesser, Singer & Trabasso, 1994; Levine, Guzmán & Klin, 2000). Such a novel-referent inference has been shown to incur a processing cost (e.g., Benatar & Clifton, 2014; Burkhardt, 2006, 2007; Burkhardt & Roehm, 2007; Murphy, 1984; Schumacher & Hung,

[1] Another possibility is that the pronoun is interpreted as referring to someone else irrespective of its gender, in which case no processing differences for matching and mismatching pronouns are predicted. This option seems unlikely given that a prominent referent is readily available for a matching pronoun.

2012; Yekovich & Walker, 1978). New referents may increase working memory load and/or representational complexity (e.g., Martin & McElree, 2008, 2011; Gibson, 1998; Gordon, Hendrick & Johnson, 2001; Warren & Gibson, 2002). However, without any further contextual information regarding the identity of the new referent, the pronoun is referentially ambiguous.

Extant discourse-based theories of pronoun resolution appear to allow for a novel-referent interpretation to occur, but do not explicitly allow the parser to overwrite pronoun gender if an agreement error is perceived[2]. For example, the memory-based framework of anaphor resolution posits that all antecedents with sufficient features in common with an anaphor are automatically activated (Gerrig, 2005; McKoon & Ratcliff, 1998; Myers & O'Brien, 1998). When the anaphor and antecedent do not have sufficient feature overlap, when the antecedent is outside of the focus of attention, or when there are competitors, the situation could occur that no antecedent resonates to sufficient degree. This situation triggers additional episodic retrieval processes to recover additional information that might help to infer the most plausible referent. For example, an attempt may be made to retrieve an antecedent from outside the focus of attention. A related account, the Bonding and Resolution framework (Garrod & Terras, 2000; Sanford & Garrod, 1989), distinguishes between the initial localization of an antecedent and the ultimate full commitment to one particular interpretation. In "the aunt found out that he" the pronoun matches the antecedent in number but not in gender, which might be sufficient for bonding to occur or for the antecedent to

[2] Comprehenders need not consistently favour one interpretation. Sentence context may bias readers away from or towards one specific referent (e.g., Nieuwland & Van Berkum, 2006; Koornneef & Sanders, 2012; Koornneef & Van Berkum, 2006). Readers are also known to differ in pronoun comprehension strategy (Almor, Kempler, MacDonald, Andersen & Tyler, 1999; Arnold, submitted; Nieuwland & Van Berkum, 2006). Additionally, readers can switch from one interpretation to another when the unfolding sentence makes a particular referential interpretation more plausible than another.

become activated, but not enough for resolution to proceed smoothly. Similarly, in a framework for pronoun comprehension based on Discourse Representation Theory (Kamp & Reyle, 1993), Gordon and Hendrick (1998) posit that a novel discourse referent is introduced after failing to find a suitable antecedent based on grammatical features such as gender, number, animacy and reflexivity. Importantly, these different accounts of pronoun comprehension clearly imply some form of processing cost when a novel referent is assumed.

The current set of studies provides data from event-related potentials (ERPs) to tease apart the two different interpretations for pronouns without a suitable antecedent. ERPs provide quantitative and qualitative information about cognitive processes via the measurement of component amplitude and component classification according to polarity (negative/positive voltage), timing of the effect onset or maximum (onset or peak latency), waveform morphology (slowly rising and sustained or peaked), and scalp distribution. Previous literature on pronoun resolution has associated two distinct ERP profiles with the detection of (or recovery from) syntactic anomaly versus referential ambiguity (Nieuwland & Van Berkum, 2006; Osterhout & Mobley, 1995).

Mismatching pronouns and Event-Related Potentials

In a canonical ERP study on agreement effects by Osterhout and Mobley (1995), participants judged the acceptability of sentences such as “The aunt heard that he won the lottery”. Mismatching pronouns elicited a P600 effect across all 12 participants, but this effect was restricted to 8 participants who tended to judge mismatch sentences as unacceptable (93% of the trials on average). No robust differential effect was obtained for the 4 participants who tended to judge mismatch sentences as acceptable (63% of the trials), although a non-significant sustained negative shift was visible for mismatching pronouns. Mismatch effects were also apparent at sentence-final words, with unacceptable-responders

showing N400 effects and no clear effects for acceptable-responders, suggesting continued processing consequences of anomaly detection on sentence comprehension. Based on the association between sentence judgments and the P600 effect, which is also observed for outright syntactic anomalies, Osterhout and Mobley concluded that mismatching pronouns were generally perceived as syntactically anomalous.

Subsequent studies suggest that something similar can happen in absence of acceptability judgment instructions (e.g, Hammer, Jansma, Lamers & Münte, 2008; Nieuwland & Van Berkum, 2006; Van Berkum, Koornneef, Otten & Nieuwland, 2007). Hammer and colleagues reported P600 effects for pronouns that mismatched the biological and also the grammatical gender of the antecedent (e.g., "Das Bübchen-$_{male}$ will schlafen und darum schaltet sie-$_{fem}$ …", approximate translation: The little boy wants to sleep and therefore she turns). Nieuwland and Van Berkum reported P600 effects when pronouns mismatched two antecedents (e.g., "David told John that she …"). Van Berkum and colleagues reported P600 effects for pronouns that matched one antecedent but mismatched the gender of the antecedent that was expected based on implicit-causality verb bias (e.g., "John hated Lisa because he …")[3]. These various P600 effects indicate that even without an acceptability judgment task, readers sometimes perceive mismatching pronouns to be syntactically anomalous. However, these studies all used different materials than Osterhout and Mobley, which complicates a direct comparison. Perception of syntactic anomaly is perhaps more likely when readers strongly expect a coreferential pronoun, which itself could be a function

[3] This P600 effect, however, had a different morphology and time course than those reported in the other studies, possibly because the correct referent was available in the sentence, and one must tread carefully in assuming similar processes underlying these effects. Nevertheless, this effect seems consistent with P600 effects observed for constructions that are well-formed but whose syntactic properties do not fit the analysis currently being pursued or that was previously expected ('garden-path' sentences; e.g., Kaan, Harris, Gibson & Holcomb, 2000; Osterhout, Holcomb & Swinney, 1994; Van Berkum, Brown & Hagoort, 1999), possibly signalling the reanalysis or recovery processes that are required to establish a meaningful interpretation (see also Osterhout & Mobley, 1997).

of the number and saliency of the antecedents, the meaning of the unfolding sentence, and the occurrence of gender mismatch at multiple levels of representation (grammatical and biological). Hence, it remains unclear whether the sentences used by Osterhout and Mobley would be perceived as syntactically anomalous when participants are not explicitly judging sentence acceptability.

Importantly, it stands to reason that a novel referent must be assumed for a mismatching pronoun to be perceived as syntactically acceptable rather than unacceptable. In discourse-based accounts of pronoun resolution (Gerrig & O'Brien, 2005; Gordon & Hendrick, 1998; McKoon & Ratcliff, 1998), pronoun gender mismatch could cause readers to try finding a suitable antecedent in the discourse context or introduce of a novel referent altogether. Both these solutions assume a referent beyond the one given in the sentence. Unfortunately, the Osterhout & Mobley results are inconclusive on this issue. Absence of a robust ERP effect in acceptable-responders could indicate that no particular processing cost was incurred for a novel referent compared to matching pronouns. However, such a conclusion seems implausible and the relevant comparison lacked sufficient statistical power. Based on discourse-based accounts of pronoun resolution, I argue that assuming a new referent assigns the problem with interpretation to the referential representation rather laying the blame on a violation of morphosyntactic agreement. The resulting prediction is that mismatching pronouns elicit an ERP effect that indexes referential ambiguity.

Previous ERP studies reliably associate referential processing difficulties with a sustained, frontal negativity (the Nref, for a review see Nieuwland & Van Berkum, 2008a). The Nref effect reflects genuine ambiguity at the level that is most relevant to discourse comprehension, the situation model (Nieuwland, Otten & Van Berkum, 2007; Nieuwland & Van Berkum, 2006; Nieuwland & Van Berkum, 2008b; Van Berkum, Brown & Hagoort, 1999). Referring expressions such as noun phrase anaphora and pronouns only elicit Nref

effects in the context of two antecedents that are equally plausible referents (Nieuwland et al., 2007; Nieuwland & Van Berkum, 2006), not when two antecedents are mentioned but one is a more plausible referent than the other. The pronoun 'he' in "John hated David because he" does not elicit an Nref presumably because comprehenders consistently take the pronoun to refer to David (as established in behavioral norms, see Nieuwland & Van Berkum, 2006), whereas 'he' in "John saw David because he" elicits an Nref as comprehenders are as likely to take John and David as referents. Moreover, Martin, Nieuwland and Carreiras (2012, 2014) reported an Nref for grammatically incorrect noun phrase determiners (the Spanish *otro/otra*, 'another', which introduces a new referent by way of ellipsis that agrees with the antecedent), which was modulated by gender of a local attractor noun. Referential processing costs can thus be incurred by an expression that could alternatively be construed as a purely morphosyntactic problem.

The current prediction that mismatching pronouns incur referential ambiguity, as indexed by the Nref, is not trivial. The difficulty with retrieving semantic information for the pronoun, or the semantic unexpectedness of the pronoun may lead to enhanced N400s (Kutas & Hillyard, 1980), as suggested by several studies. King and Kutas (1997; unpublished data, discussed in Kutas, Federmeier, Coulson, King & Munte, 2000) reported an N400 effect (plus sustained negativity) for pronouns that mismatched stereotypical gender nouns (e.g., "The engineer redesigned the circuit because she …"), which could reflect difficulty with establishing coreference with a stereotypical gender nouns. Interpretation of their findings was complicated by enlarged N400s for feminine pronouns compared to masculine pronouns irrespective of antecedent gender-stereotype. Other studies reported N400 gender-mismatch effects for smaller antecedent-pronoun distances but P600 effects for larger distances (Qiu et al., 2012; Hammer et al., 2008), and concluded that working memory therefore modulates the role of semantic information in pronoun comprehension. N400 effects have also been

reported for pronouns with an unfocused referent set (e.g., 'their' following "Not many of the fans", as opposed to following "Many of the fans"; Filik, Leuthold, Moxey & Sanford, 2011), and for proper names when coreference is infelicitous (e.g., "John said that John", Swaab, Camblin & Gordon, 2004). In sum, mismatching pronouns may incur semantic retrieval difficulties as indexed by N400 modulations, but there is nothing to suggest that these difficulties arise from assuming a new referent or from not knowing who the intended referent is. In contrast, referential ambiguity is reliably indexed by the Nref effect (Nieuwland & Van Berkum, 2008a).

Experiment 1 examined the processing consequences of mismatching pronouns in sentences that were very similar to those used by Osterhout and Mobley (1995). But instead of a sentence acceptability task that potentially biases participants towards treating mismatching pronouns as unacceptable, task instructions promoted the addition of a novel referent for a mismatching pronoun. Under these task instructions, are mismatching pronouns still processed as if they were an agreement error or do they incur a referential processing cost instead? Experiment 2 and 3 investigate the processing consequences of gender mismatch without any task instructions regarding the pronouns.

EXPERIMENT 1

Experiment 1 investigated the ERP processing profile of gender-mismatching pronouns under a task instruction that discouraged the detection of syntactic anomaly and encouraged an extra-sentential referent interpretation. Participants were asked to read sentences and try and infer a new referent for mismatching pronouns (see Table 1 for example sentences, a full set of materials is available as an appendix). Participants were instructed to think of a new referent that would fit the described scenario most plausibly, but were not instructed about who this referent should be (other than that it should be a novel

referent for each sentence). This particular instruction, although untested, was chosen rather than an acceptability judgment task, as the latter may bias participants towards judging mismatching pronouns as unacceptable (which could leave insufficient participants for the relevant comparison, see Osterhout & Mobley, 1995). The new-referent instruction was intended to tap into the relevant processes more directly than instructing participants to judge mismatching pronouns as 'acceptable' in an acceptability judgment task. The prediction for the pronoun mismatch sentences was that under these task constraints, an Nref would be elicited rather than a P600 effect, as the Nref indexes referential ambiguity (Nieuwland & Van Berkum, 2008a; Van Berkum et al., 2007). For comparison, fillers sentences with 'straightforward' semantic and syntactic anomalies were used to elicit standard N400 and P600 effects (based on materials of Osterhout & Nicol, 1999).

An additional comparison was made between sentence-final words in sentences with matching and mismatching pronouns. Osterhout and Mobley reported that sentence-final words in unacceptable sentences elicit sentence-final N400 effects, possibly because an acceptable interpretation has not yet been derived by that time (see also Osterhout & Holcomb, 1992; Osterhout et al., 1997a). Consistent with this idea, sentences that are judged as acceptable do not incur sentence-final enhanced N400s. Thus, if despite the new-referent instructions participants perceive mismatching to be syntactically anomalous, as indexed by a P600 effect, an accompanying sentence-final N400 effect is predicted.

Experiment 1 addressed two additional issues that go beyond a replication of the Osterhout & Mobley study with different task instructions. The first one was the potential role of working memory processes in the comprehension of mismatching pronouns. Assuming a new referent may tax online processing resources more than sentences with matching pronouns (for related findings from computational modelling, see Van Rij, Van

Rijn & Hendriks, 2011). Individual differences in ERP responses might then be expected to correlate with a standard measure of working memory functioning such as Reading Span (Daneman & Carpenter, 1980). The Reading Span test requires participants to read aloud sets of sentences while remembering the sentence-final words, tapping into the capacity to process and store verbal information simultaneously. High span readers generally perform better on both off-line and on-line measures of language comprehension (Just and Carpenter, 1992), and they are more likely to elaborate their discourse models with optional, knowledge-based inferences than low span readers (e.g., St. George, Mannes & Hoffmann, 1997). In previous ERP work on pronoun comprehension (Nieuwland & Van Berkum, 2006), high span readers showed Nref effects for ambiguous pronouns whereas low span readers did not, suggesting that the former group indeed noticed referential ambiguity whereas the latter group immediately took on the most readily available referential interpretation. In the current study, low span readers might thus be more likely to interpret the pronoun coreferentially, and thus show stronger P600 activity. In contrast, high span readers may be more prone to assume a new referent, and as this would incur referential ambiguity, they would show stronger Nref activity.

Experiment 1 also explored potential differences in pronoun resolution between gender-definitional noun phrase antecedents (e.g., "the businessman") and proper name antecedents (e.g., "John"), while Osterhout and Mobley used noun phrase antecedents only. Naming is a major factor in discourse focus control (Sanford, Moar & Garrod, 1988), as named characters are more likely to be story protagonists, whereas noun phrase role descriptors tend to be secondary characters. Named characters are more prominent and therefore more accessible to readers, as suggested by faster reading times and probe recognition times for pronominal reference to named characters compared to role descriptors (e.g., Garrod, Freudenthal & Boyle, 1994; Mcdonald & Shaibe, 2002; Sanford et al., 1988).

In Experiment 1, this difference in representation strength (see also Foraker & McElree, 2007) may impact the likelihood that a coreferential interpretation is attempted. This could make P600 effects of mismatch more likely to occur for named antecedents than for noun phrase antecedents, while perhaps Nref effects more likely to occur for noun phrase antecedents than for named antecedents.

Table 1. Example sentences for Experiment 1 and 2. The full list of materials is available in the appendix.

Pronoun sentences: Match, Mismatch

1. The boy thought that he/she would win the race.
2. John shouted that he/she was very angry today.
3. The businessman observed that he/she was very successful recently.
4. Clifford mentioned that he/she was getting a divorce.
5. The aunt intimated that she/he was not happily married.
6. Anne presumed that she/he would ace the exam.
7. The choirgirl hoped that she/he would pass the audition.
8. Cassandra was surprised that she/he enjoyed the terrible movie.

Filler sentences: Correct, Syntactic anomaly, Semantic anomaly

1. A new computer will last/lasting/paint for many years.
2. Simple vegetable oil is used to fry/frying/plow the vegetables.
3. The alarm system should warn/warning/swear that there is an intruder.

METHODS

Participants

Twenty-six right-handed University of Edinburgh students (10 men, between 18 and 35 years old) gave written consent. All were native English speakers without neurological or psychiatric disorders and were paid for their participation.

Materials

One hundred and sixty sentence pairs were constructed, 80 starting with a male or female proper name and 80 starting with the determiner 'the' followed by a male or female gender-definitional noun (partly based on materials published by Osterhout et al., 1997a; Osterhout and Mobley, 1995). Gender-definitional nouns were chosen because they tend to elicit more robust gender-mismatch effects than gender-stereotypical nouns (Kreiner et al., 2008; Osterhout et al., 1997a). These sentential subjects were always followed by 'verb-ed that' and the gender-matching or -mismatching pronoun 'he' or 'she' and 4 further words.

Fillers consisted of 120 sentence triplets that contained critical verbs that could be syntactically anomalous, semantically anomalous, or grammatically and semantically unproblematic (e.g., "The beavers sometimes chewing/melt/chew"; partly based on materials published in Osterhout & Nicol, 1999). Critical verbs in filler sentences were followed by at least two additional words. A further 40 grammatically and semantically unproblematic sentences were added as fillers so that half of the total set of sentences was unproblematic.

Six lists were created so that each of the 320 sentences appeared in only one condition per list, but in all conditions equally often across lists. Each participant thus saw a total of 80 gender-matching and 80 gender-mismatching pronoun sentences, 40 syntactically anomalous sentences, 40 semantically anomalous sentences, and 80 unproblematic filler sentences.

Sentence types within each list were pseudo-randomly mixed to limit the repetition of conditions.

Procedure

Participants read sentences from a monitor (black letters, light-gray background), presented word-by-word at a regular pace (300 ms word duration, 200 ms inter-word-interval) with sentence-final words presented for 1000 ms, followed by a fixation cross upon which participants could self-pace to the next sentence by button-press. Participants were informed that the sentences could contain pronouns with a gender that did not match the person in the sentence, and were instructed to think of a new person whom the pronoun referred to. No specific instruction was given as to whom this new person should be, but participants were asked to think of a new person for each new sentence that would render the described scenario involving two people plausible as a whole. Post-experiment feedback from participants about this task included that sometimes they waited until the end of the sentence to infer the new referent, and that the task was intensive yet engaging and got easier throughout the experiment. Participants completed a practice-session and eight break-separated experimental sessions. Total time-on-task was 50 min.

After the ERP experiment, participants performed a computerized Reading Span test (Van den Noort, Bosch, Haverkort & Hugdahl, 2008; see also Nieuwland & Van Berkum, 2006). Reading Span score was computed as the total number of words that were correctly recalled (Maximum = 100).

Electroencephalogram recording and data processing

The electroencephalogram *(*EEG) was recorded at a sampling rate of 512 Hz using a BioSemi ActiveTwo system (http://www.biosemi.com) with 64 EEG electrodes in an

international 10–20 electrode configuration (see Figure 1), two additional mastoid electrodes and four EOG electrodes (left and right horizontal cantus, and above/below the right eye), referenced to the common mode sense (CMS; active electrode) and grounded to a passive electrode. The EEG was re-referenced to the average of the left and right mastoid electrode offline, filtered (0.019-20 Hz band-width filter plus 50 Hz Notch filter) and corrected for ocular artefacts using independent component analysis as implemented in BrainVision Analyzer (Brain Products ©). Data was then segmented into epochs that started 200 ms before word onset, and that lasted until 1500 ms after pronoun onset or until 1000 ms after word onset for filler conditions and sentence-final words. All epochs were baseline-corrected and then automatically screened for artefacts (minimal/maximal allowed amplitude = -75/75 μV) before entering into condition-averages per participant.

Two participants were excluded due to excessive artefacts. Cut-off was 40 artefact-free pronoun epochs and sentence-final sentences and 20 artefact-free trials for filler sentences. For the remaining 24 participants, average ERPs were computed over artefact-free trials for pronouns (match, M = 65, *S.D.* = 9, mismatch, M = 63, *S.D.* = 11), sentence-final words (match, M = 70, *S.D.* = 9, mismatch, M = 70, *S.D.* = 8), and filler sentences (control, M = 36, *S.D.* = 3, semantic violation, M = 36, *S.D.* = 3, syntactic violation, M = 36, *S.D.* = 3).

Statistical analysis

Using average amplitude per condition across all EEG electrodes, a 2(Gender-match: match, mismatch) repeated measures analysis of variance (ANOVA) was performed in consecutive time windows based on a previously published ERP investigation on the functional nature of the Nref (Nieuwland et al., 2007): 100-300, 300-600, 600-1000 and 1000-1500 ms. Statistical analysis of the filler sentences as well as for sentence-final word

data focused on the 300-500 and the 500-800 ms time window to test for N400 and P600 effects respectively, following Osterhout and Mobley (1995).

Scalp distributions of the observed effects were examined using electrode grouping into Regions-Of-Interest (ROIs, see Figure 1). Separate analyses were performed for Lateral ROIs (LAF/RAF, LLFC/RLFC, LLCP/RLCP, LPO/RPO) using a 2(Gender-match: match, mismatch) by 2(Hemisphere: left, right) by 4(Anteriority: Anterior-Frontal, Frontal-Central, Central-Parietal, Parietal-Occipital) ANOVA, for Medial ROIs (LMFC/RMFC, LMCP/RMCP) using a 2(Gender-match: match, mismatch) by 2(Hemisphere: left, right) by 2(Anteriority: Frontal-Central, Central-Parietal) ANOVA, midline ROIs (MAF/MFC/MCP/MPO) using a 2(Gender-match: match, mismatch) by 2(Anteriority: Anterior-Frontal, Frontal-Central, Central-Parietal, Parietal-Occipital) ANOVA, and crossline ROIs (LLC/LMC/RMC/RLC) using a 2(Gender-match: match, mismatch) by 2(Hemisphere: Left-Lateral, Left-Medial, Right-Medial, Right-Lateral) ANOVA. Where appropriate, Greenhouse–Geisser corrections and corrected F-values are reported. Only statistical results with $p < .1$ are reported.

RESULTS

As shown in Figure 2 (upper pane), mismatching pronouns elicited a left-frontally distributed, sustained negative shift started at about 300 ms and lasted throughout the whole epoch, and that was most prominent at electrode F3. In addition, mismatching pronouns elicited a more positive ERP component at left-posterior channels between approximately 500 and 1000 ms (P600 effect). At the sentence-final words (lower-left pane), the mismatch condition elicited larger N400s than match conditions, with the N400 effect being centrally distributed. In the filler sentences (lower-right pane), semantic violations elicited an N400 effect compared to the other two conditions, and syntactic violations elicited a large P600

effect, both these effects had a broad distribution across the scalp. Statistical results for the filler sentences are available as supplementary materials. Supplementary figures that show all electrodes for the pronoun data and sentence-final word data are available on the JML website.

Pronoun results

300-600 ms: A robust gender-match by anteriority by hemisphere 3-way interaction effect was observed in the lateral analysis ($F_{3,69} = 3.6$, $p < .05$) and in the medial analysis ($F_{1,23} = 4.4$, $p < .05$). Lateral follow-up analysis revealed that ERPs elicited by mismatch were marginally significantly more negative in the LLFC-ROI (M= -.56, $S.E.$ = .31, $F_{1,23}$ = 3.2, p = .085), and the medial follow-up analysis showed this pattern for the LMFC-ROI (M= -.88, $S.E.$ = .45, $F_{1,23} = 3.8$, p = .06).

600-1000 ms: A robust gender-match by anteriority by hemisphere 3-way interaction effect was observed in the lateral analysis ($F_{3,69} = 5.2$, $p < .01$) and in the medial analysis ($F_{1,23} = 6.5$, $p < .05$). Lateral follow-up analysis revealed that mismatch elicited more positive ERPs in the LPO-ROI (M= 1.1, $S.E.$ = .45, $F_{1,23} = 5.6$, p < .05).

1000-1500 ms: As in the preceding windows, both the lateral and the medial analysis revealed a robust gender-match by anteriority by hemisphere 3-way interaction effect ($F_{3,69}$ = 6.9, p = .001; $F_{1,23} = 13.4$, p = .001, respectively). Medial follow-up analysis revealed that mismatch elicited more negative ERPs in the LMFC-ROI (M= -1.1, $S.E.$ = .52, $F_{1,23} = 5.3$, p < .05).

Reading Span results

Reading span scores ranged from 52 to 91 (M = 74, $S.D.$ = 11). Correlations and reading span median-split were performed in the 600-1000 and 1000-1500 time windows using the ROIs that had shown effects of gender-match. In the 600-1000 ms and 1000-1500 ms time window, higher reading span was associated with more negative ERPs at the anterior channels (600-1000: LLFC-ROI: r = -.44, p < .05; LMFC-ROI: r = -.44, p < .05, 1000-1500: LLFC-ROI: r = -.42, p < .05; LMFC-ROI: r = -.38, p = .07), and to a lesser extent at the posterior ROI (600-1000: LPO-ROI: r = -.38, p = .07).

Median split analyses confirmed that high span participants showed larger Nref effects than low span individuals ($F_{1,22}$ = 7.5, p < .05), and that the Nref effect was robust in high span participants (M= -1.8, $S.E.$ = .61, p < .01) but not in low span participants (M= .36, $S.E.$ = .51, ns.). The reverse was true for the P600 effect, although the group by mismatch interaction effect was marginally significant ($F_{1,22}$ = 3.8, p < .1), follow-up showed a robust P600 effect in low span individuals (M= 1.77, $S.E.$ = .56, p < .005) but not in high span individuals (M= .1, $S.E.$ = .66, ns.).

Sentence-final word data

300-500 ms: Sentence-final words following mismatching pronouns elicited larger (more negative) N400s across all electrodes (M= -.82, $S.E.$ = .31, $F_{1,23}$ = 7.1, p < .05), with no robust distributional effects except a gender-match by hemisphere interaction effect in the crossline analysis ($F_{3,69}$ = 3.9, p < .05): this N400 effect was statistically reliable in the LLC, LMC and RMC ROIs (all Fs > 7.7) but not in the RLC-ROI.

500-800 ms: Sentence-final words following mismatching pronouns did not elicit robust differences, although a reliable gender-match by hemisphere interaction effect in the

crossline analysis ($F_{3,69} = 4.6$, $p < .01$) suggested that the enhanced negativity for mismatch was somewhat prolonged in the LLC-ROI (M= -.99, $S.E.$ = .42, $F_{1,23} = 5.6$, $p < .05$).

An additional analysis was conducted to relate ERP responses to pronouns with those to sentence-final words. In the Osterhout and Mobley (1995) study, individuals who responded 'acceptable' had not shown any reliable effect (although this group was probably too small to generate statistically robust effects), whereas individuals who responded 'unacceptable' generated a P600 effect plus a sentence-final N400 effect. Here, an additional correlation test examined whether the P600 modulation at the LPO-ROI in the 600-1000 ms time window predicted the N400 modulation at the sentence-final words (average value across medial ROIs): larger pronoun-elicited P600 effects were indeed associated with larger sentence-final N400 effects (r = -.47, $p < .05$). This correlation appeared not simply due to variability in overall size of ERP responses: correlation between pronoun-elicited P600 effects and syntactic violation-elicited P600 effects (500-800 ms time window at the LPO-ROI; see analyses below) was close to zero (r = -.03, *n.s.*).

As low span readers showed larger pronoun-elicited P600 effects, they were also the group that showed larger sentence-final N400 effects. Correlation between reading span and sentence final N400 effect (all electrodes, 300-500 ms time window) was only marginally significant (r = -.37, $p < .01$), but median-split analysis showed a significant reading span group by mismatch effect ($F_{1,22} = 5.1$, $p < .05$), with a robust sentence-final effect in low-span individuals (M= -1.36, $S.E.$ = .37, p = .001) but not in high-span individuals M= -.06, $S.E.$ = .44, ns.).

Proper name antecedents versus noun phrase antecedents

In this exploratory analysis, epochs in the pronoun match and mismatch conditions were split based on antecedent type (see Figure 3), and statistical analysis was performed using a 2(Gender-match: match, mismatch) x 2(Antecedent: proper name, noun phrase) ANOVA. Based on the results in the main analysis, testing was performed at the LPO-ROI and the combined LLFC/LMFC-ROI.

At the LPO-ROI, a significant gender-match by antecedent interaction was found in the 300-600 ms time window ($F_{1,23} = 4.3$, $p < .05$), in which proper name mismatch elicited more positive ERPs than match ($M = 1.2$, *S.E.* = .46, $F_{1,23} = 6.2$ $p < .05$), but no difference was found for noun phrase mismatch ($M = .04$, *S.E.* = .43, *ns*.). No robust interaction effect was observed in the 600-1000 ms time window.

At the combined LLFC/LMFC-ROI, the full 300-1500 ms time window was used based on the sustained nature of the Nref modulation. A marginally significant gender-match by antecedent interaction was found ($F_{1,23} = 3.1$, $p < .1$), in which noun phrase mismatch elicited more negative ERPs than match ($M = -1.05$, *S.E.* = .46, $F_{1,23} = 5.2$, $p < .05$), whereas no such effect occurred for proper name mismatch ($M = -.09$, *S.E.* = .45, *ns*.).

A post-hoc survey was conducted to investigate whether people found it easier to think of a new referent for noun phrase sentences or for proper name sentences. Twelve native speakers of English read each sentence truncated after the mismatching pronoun. They were told that the pronouns referred to someone new and were asked to judge how easily this someone came to mind (1 = difficult, 3 = easy). Participants judged sentences with noun phrases to be somewhat easier in this regard ($M = 2.37$, *S.D.* = .26) than with proper names ($M = 2.29$, *S.D.* = .23), although this small difference was marginally significant ($t(158) = 1.94$, $p = .054$, two-tailed).

Analyses involving anaphoric pronoun bias

Another exploratory analysis was performed to examine whether the P600 modulation for mismatching pronouns depended on the expectation of a matching pronoun (e.g., Van Berkum et al., 2007). Twenty-nine native speakers of English who had not participated in the ERP experiment completed all sentences, truncated before the pronoun, with the first plausible completion that came to mind. An 'anaphoric bias score' was computed for each sentence by subtracting the number of completions that started with a new referent (e.g., mismatching pronouns, proper names, noun phrases) from the number of completions that started with a gender-matching pronoun (other completions were not counted). A high/positive anaphoric bias score thus indicates that the sentence was commonly expected to continue with information about the subject (e.g., "*The schoolgirl whispered that*"), whereas a low/negative anaphoric bias score indicates that the sentence was commonly expected to continue about a referent other than the subject (e.g., "*The father hoped that*"). The items had an overall positive anaphoric bias (mean $M = 7.4$, $S.D. = 10.4$, median = 7). Although average anaphoric bias was nominally higher for proper name sentences ($M = 8.7$, $S.D. = 9.4$, median = 7) than for noun phrase sentences ($M = 6.2$, $S.D. = 11.2$, median = 6), this difference was not significant ($t(158) = 1.54$, n.s.). Trials in match and mismatch pronoun conditions were split into relatively low and high anaphoric bias based on the median score of the trials (per condition, per subject) that had gone into the subject average (see Figure 4). Statistical analysis was then performed in the 600-1000 ms time window using with a 2(Gender-match: match, mismatch) x 2(Anaphoric bias: low, high) design at the LPO-ROI. Although no significant interaction was observed in this time window ($F < 1$), post-hoc pair-wise comparisons showed that the mismatching pronouns elicited a robust P600 effect in the high-bias condition ($M= 1.6$, $S.E. = .68$, $p < .05$), but not in the low-bias condition ($M= .80$, $S.E. = .62$, n.s.).

DISCUSSION

Experiment 1 examined the electrophysiological responses to pronouns that mismatched or matched gender of the only available antecedent. At the group-level, mismatching pronouns elicited two distinct effects: (1) a frontal, left-lateralized sustained negativity that started about 300 ms after pronoun onset and was strongest in the 300-600 and 1000-1500 time windows, (2) a left-posterior P600 effect that started at about 500 ms and that was robust in the 600-1000 ms time window. The waveform morphology, frontal scalp distribution and timing of the sustained negative ERP response are consistent with previous Nref effects for ambiguous pronouns or noun phrases (Nieuwland & Van Berkum, 2006). This effect lacked statistical significance in the 600-1000 ms window, however, which could be explained by overlap from the observed P600 effect, as negative and positive voltage cancel each other out at the scalp surface. In turn, the P600 modulation may have been counteracted by Nref overlap. The observed P600 effect had a very limited distribution but was otherwise very similar in onset and waveform morphology to P600 effects reported by Osterhout and Mobley and by Nieuwland and Van Berkum (2006), and also to the P600 effect observed for syntactic violation filler sentences in the current study. Component overlap of the Nref and P600 is likely due to the extended spatial distributions of both these effects (e.g., Nieuwland & Van Berkum, 2006), even as these effects are usually maximal at different locations.

Most importantly, the observation of two distinct ERP effects is not a straightforward replication of the Osterhout and Mobley P600 effect findings. However, like those findings, the current mixture of results seemed to involve individual differences in ERP profiles. In Osterhout & Mobley, ERP responses were a function of the tendency to judge sentences mismatching pronoun as acceptable or unacceptable. Here, individuals with high reading

span scores showed larger Nref effects and smaller P600 effects, whereas individuals with lower reading span showed larger P600 effects and smaller Nref effects. This finding suggests that high span individuals were more likely to assume a novel referent for the mismatching pronouns, whereas low span individual were more likely to attempt, at least initially, a coreferential interpretation. This is broadly consistent with the conclusion reached by Nieuwland and Van Berkum (2006) that high span individuals are more sensitive to or more likely to entertain alternative referential interpretations, whereas low span individuals are inclined towards the less costly, readily available interpretation (see also MacDonald & Christiansen, 2002). There could be an interesting parallel here with Osterhout et al. (1997a), who investigated the comprehension of reflexive pronouns that matched or mismatched the stereotypical gender of the antecedent (e.g., "The electrician shocked himself/herself"). Mismatching sentences that were judged as acceptable nevertheless elicited a P600 effect as also seen to sentences judged as unacceptable. Osterhout et al. concluded that participants initially assigned the 'preferred', stereotypical gender of the antecedent, but revised the gender following the mismatching reflexive. Here, under the novel-referent task instruction, the P600 effect suggests that some participants, particularly low reading span individuals, attempted a coreferential interpretation first and subsequently had to overcome that initial interpretation in response to the instructions. Of note, the P600 effect thus need not directly reflect the detection of a syntactic anomaly, but could reflect the subsequent recovery processes involved in deriving a coherent grammatical analysis (Osterhout et al., 1994, 1997a).

The exploratory analyses of antecedent type and anaphoric bias suggest that the observed ERP effects may also be modulated by sentential factors, although none of the tested interaction effects were statistically robust. Across all participants, noun phrase antecedents were associated with an Nref effect, whereas proper name antecedents were

associated with a different positive ERP effect. One speculative interpretation of the latter effect is that it is a P600 effect because participants are more likely to attempt a coreferential interpretation for proper name antecedents because they are more prominent than noun phrase antecedents (e.g., Sanford et al., 1988). Related to this point, many of the noun phrase antecedents used in the current item set already entailed another referent by virtue of a category relationship (e.g., 'mother' entails a child), which may have made it less likely that participants attempted to establish coreference. One very important caveat to this interpretation, however, is that the differential effect for proper name antecedents was not robust in the 600-1000 ms window, but in the 300-600 ms time window and was perhaps not a P600 effect. Based on the time course and on the posterior scalp distribution (see Figure 3), this differential effect is more consistent with an enhanced N400 component for matching pronouns rather than an enhanced P600 for mismatching pronouns. Such an N400 effect of antecedent type has not been reported previously. Further research is needed, using greater trial numbers, to provide stronger support for an antecedent type effect in processing mismatching pronouns, and must control for other factors such as anaphoric bias and antecedent frequency. In contrast to the effects of antecedent type, the effect of anaphoric bias seemed more clearly concentrated on the P600 modulation with the differential effect being driven by the mismatch conditions rather than by the matching conditions as was the case in the antecedent type effects. Although the anaphoric bias by match interaction effect was not statistically robust[4], pairwise tests showed that the P600 effect reached significance in the high-bias items and not in the low-bias items. The results of this exploratory analysis

[4] A potential reason that the match by anaphoric interaction test was not significant was because it was not as strong as where the main P600 effect was observed (at the LPO-ROI) and where the interaction was tested (for example, P3 showed the strongest difference in P600 effect modulation by anaphoric bias but lies outside the LOP-ROI), and the 600-1000 time window may also not be optimal to pick up on P600 differences compared to a more limited time window where the P600 modulation is strongest (e.g., 600-800 ms).

must thus also be treated with caution, but they are generally consistent with the prediction that people are more likely to attempt a coreferential interpretation for a mismatching pronoun when they strongly expected a matching pronoun (for relevant discussion on choice of referring expressions, see Fukumura &Van Gompel, 2010; Fukumura, Van Gompel, Harley & Pickering, 2011; Fukumura, Hyönä & Scholfield, 2013; see also Kehler & Rohde 2013; Rohde & Kehler, in press).

Given the lack of a behavioral task that directly linked the observed P600 effect to a perceived anomaly, an alternative explanation might be that the observed P600 effect reflects the addition of a novel referent. Such has indeed been claimed for the introduction of new referents by comparing novel versus given noun phrases (e.g., Burkhardt, 2006, 2007; Schumacher & Hung, 2012; see also Kaan, Dallas & Barkley, 2007), although there is no evidence to suggest that this is the case for pronouns. I will return to this issue in the General Discussion, and provide several reasons for why this alternative explanation seems implausible given the literature on pronoun comprehension and given the currently observed results.

One potential concern about Experiment 1 is the nature of the new-referent task. I wish to emphasize that this task was not the main focus of the experiment but was solely used to discourage participants from treating the mismatching pronouns as syntactic anomalies. However, although most participants said that they found the task engaging, adherence to the instruction can only be assumed because there was no monitoring or quantification of actual performance. And even if participants indeed introduced a novel referent for each mismatching pronoun, there is no certainty regarding the time at which they did. They could have done so while reading or possibly have waited until the end of each sentence. Moreover, the instruction of the task explicitly pointed out to participants that there could be a mismatch

of the pronoun gender and the antecedent. This may have focused participants' attention on the task-relevant grammatical features of the pronoun, similar to how participants pay attention to pronoun gender in an acceptability judgment task. In sum, the contribution of the task instructions to the observed results is unclear. Experiment 2 was performed to address this issue, and was identical to Experiment 1 except that it omitted the new-referent instruction.

EXPERIMENT 2

METHODS

All methods were identical to those of Experiment 1 except for the participants and the instruction. Participants in Experiment 2 were instructed to read the sentences for comprehension but nothing was said about the different types of sentences nor was any instruction given as to how they should interpret the sentences.

Participants

Twenty-two right-handed University of Edinburgh students (8 men, between 18 and 35 years old) gave written consent. All were native English speakers without neurological or psychiatric disorder, and had not participated in Experiment 1.

Three participants were excluded due to excessive artefacts. For the remaining 19 participants, average ERPs were computed over artefact-free trials for pronouns (match, M = 63, $S.D.$ = 12, mismatch, M = 63, $S.D.$ = 13), sentence-final words (match, M = 70, $S.D.$ = 8, mismatch, M = 68, $S.D.$ = 8), and filler sentences (control, M = 36, $S.D.$ = 3, semantic violation, M = 36, $S.D.$ = 4, syntactic violation, M = 36, $S.D.$ = 4).

RESULTS

As shown in Figure 5 (upper pane), mismatching pronouns elicited a left-frontally distributed, sustained negative shift started at about 300 ms and lasted throughout the whole epoch, and that was most prominent at electrode F5, F7 and AF7. There was no sign of the P600 effect that was observed in Experiment 1. At the sentence-final words (lower-left pane), the mismatch condition also did not elicit the clear N400 modulation that was seen in Experiment 1. In the filler sentences (lower-right pane), semantic violations elicited a small N400 effect compared to the other two conditions, whereas syntactic violations elicited a large P600 effect. Both these effects had a broad distribution across the scalp, although they both were stronger at posterior channels. Statistical results for the filler sentences are available as supplementary materials.

Pronoun results

300-600 ms: Mismatching pronouns elicited more negative ERPs across all electrodes although this effects was only marginally significant (M = -.57, $S.E.$ = .28, $F_{1,19}$ = 4.18, p = .056). In the lateral analysis this match effect was significant ($F_{1,19}$ = 4.7, p < .05), but no robust interactions with anteriority or hemisphere were found. In the medial, midline and crossline analysis marginally significant effects of match were found (all ps < .1) but no significant interaction with distribution.

600-1000: The medial analysis revealed a statistically significant gender-match by anteriority interaction effect ($F_{1,18}$ = 4.6, p < .05): Mismatching pronouns elicited more negative ERPs in the LMFC-ROI ($F_{1,18}$ = 7.8, p < .05).

1000-1500: Mismatching pronouns elicited more negative ERPs across all electrodes although this effect was marginally significant (M = -.56, $S.E.$ = .31, $F_{1,19}$ = 3.2, p = .09). The

lateral analysis also showed that mismatching pronouns elicited more negative ERPs ($F_{1,19}$ = 5.6, $p < .05$) but there were no significant distributional effects.

Reading Span results

Reading span scores ranged from 48 to 81 (M = 68, $S.D.$ = 9). Correlation analysis was performed between reading span score and the robust gender match effect at the LMFC-ROI in the 600-1000 time window. Consistent with the findings from Experiment 1, the negativity to mismatching pronouns was larger for individuals with higher reading span scores (r = -.47, $p < .05$).

Median split analyses revealed only a marginally significant group effect ($F_{1,17}$ = 3.2, $p < .1$), but as in Experiment 1 the Nref effect was robust in high span participants (M= -1.3, $S.E.$ = .38, $p < .005$) but not in low span participants (M= -.33, $S.E.$ = .36, ns.).

Sentence-final word data

300-500 ms: The lateral analysis revealed a gender-match by anteriority by hemisphere 3-way interaction effect ($F_{3,54}$ = 4.8, p = .01), and the medial analysis revealed a gender-match by hemisphere effect $F_{1,18}$ = 9.6, $p < .01$), but none of the follow-up tests revealed significant effects. No impact of reading score on the sentence-final ERPs was found.

500-800 ms: No statistically significant effects were found in this time window.

Because mismatching pronouns appeared to elicit a more positive P200 component, an additional analysis was performed in the 200-300 ms time window, but this analysis also did not generate any robust results.

Proper name antecedents versus noun phrase antecedents

Grand-average ERPs for matching and mismatching pronouns following proper name and noun phrase antecedents are shown in Figure 3. Because of the sustained nature of the observed Nref effect, testing was performed using mean values from the 300-1500 ms time window at the combined LLFC/LMFC-ROI. The mismatch by antecedent interaction effect was marginally significant ($F_{1,18} = 3.2$, $p < .1$), with the pairwise comparison showing a strong Nref mismatch effect for proper name mismatch ($M = -1.3$, $S.E. = .34$, $F_{1,19} = 14.9$, $p = .001$) but no such effect for noun phrase mismatch ($M = -.35$, $S.E. = .44$).

Anaphoric bias

Grand-average ERPs for matching and mismatching pronouns in high and low anaphoric bias conditions are shown in Figure 4. Although the mismatch effect appeared to be more widespread in the high anaphoric bias condition, no significant interaction between anaphoric bias and gender mismatch was found (300-1500 ms time window, LLFC/LMFC-ROI; $F < 1$, n.s.).

DISCUSSION

In contrast to Experiment 1, mismatching pronouns only elicited a sustained frontal negativity (Nref). This Nref had a more widespread scalp distribution than the effect in Experiment 1 (especially in the 1000-1500 ms time window, while it was more left-lateralized in the 600-1000 window), and it was also statistically significant in the 600-1000 ms time window where no robust Nref effect was observed in Experiment 1. Moreover, while the filler sentences replicated the standard N400 and P600 effects observed in Experiment 1, the effects at the sentence-final words in Experiment 1 disappeared along with the pronoun-elicited P600 effect. Experiment 2 also replicated the effect of reading span on the observed

Nref effect that was seen in Experiment 1. Individuals with higher reading span scores showed a larger Nref effect than individuals with lower reading span scores. This time, however, individuals with lower reading span scores did not elicit the P600 effect seen in Experiment 1. Moreover, the effects of antecedent type and anaphoric bias were different in Experiment 2 than in Experiment 1. In Experiment 1, mismatching pronouns elicited a P600 effect following proper name antecedents, but an Nref effect following noun phrase antecedents. In Experiment 2, a robust Nref effect was only observed for proper name antecedents, although no robust antecedent type by match interaction effect was observed. No modulation of anaphoric bias was observed in Experiment 2.

The conclusions that can be drawn from the results of Experiment 2 are the following: (1) The observation of an Nref effect suggests that participants did not treat mismatching pronouns as morphosyntactic (preference) violations, as arguably would have been reflected in a P600 effect (Nieuwland & Van Berkum, 2006; Osterhout & Mobley, 1995; Van Berkum et al., 2007). This interpretation also received support from the lack of sentence-final effects. In Experiment 1, pronoun-elicited P600 effects predicted the sentence-final N400 effects (see also Osterhout & Mobley, 1995). Although conclusions based on lack of a differential effect are necessarily tentative, absence of sentence-final N400 effects in Experiment 2 suggests that a grammatically coherent analysis had been reached by the end of the sentence (see also Osterhout et al., 1997a). (2) The modulation of the Nref by reading span replicated this finding in Experiment 2, and suggests that individuals with higher reading span were more likely to assume a new referent for mismatching pronouns than individuals with lower reading span. But given that latter group did not show a clear P600 effect instead, the question arises how to interpret the lack of a robust Nref effect, or lack of any effect for that matter. One potential interpretation of these patterns is that these individuals did not attempt to establish a coreferential interpretation nor did they attempt to find a novel referent for the

mismatching pronoun. In other words, some individuals may have been processing the pronouns rather shallowly, without committing to any specific referential interpretation, and without trying to resolve ambiguioty. This links back to previous findings from Green, McKoon & Ratcliff (1992), who reported that pronouns are sometimes unresolved, even when a pronoun's referent is available and unambiguous. As argued by Love and McKoon (2011), whether pronouns are left unresolved might be a function of referent accessibility but also readers' engagement with the text. Without the novel-referent instruction, engagement with the set of unrelated sentences was likely to be less. (3) A robust Nref effect was found following proper name antecedents but not for noun phrase antecedents. This contrasts with Experiment 1, and suggests that even for more salient antecedents, a coreferential interpretation is not automatically elicited.

I will return to these issues in the General Discussion, after reporting the findings of Experiment 3. Experiment 3 was intended to solidify the main claim of this paper and replicate the results from Experiment 2 under different experimental circumstances. Because Experiment 1 and 2 contained a large number of pronoun sentences (half of the total number of sentences), the observed results may arise from strategic behaviour in response to being presented such a large number of mismatching pronouns. Thus, if participants are more likely to attempt a coreferential interpretation when there are fewer pronoun sentences, mismatching pronouns then would elicit a P600 effect. Alternatively, if participants take mismatching pronouns to be referentially ambiguous then a replication of the Nref effect is expected.

EXPERIMENT 3

This experiment attempted to replicate the findings from Experiment 2 with a smaller number of pronoun sentences. The experiment involved manipulation-orthogonal

comprehension questions to encourage participants to pay attention to the meaning of the sentences. The selected pronoun sentences were mixed with sentences from another experiment, and therefore did not contain the same filler sentences as the previous experiments, but used relatively long filler sentences that contained no semantically or syntactically problematic sentences.

METHODS

Participants

Forty-two right-handed University of Edinburgh students (16 men, between 18 and 35 years old) gave written consent. All were native English speakers without neurological or psychiatric disorder and were paid for their participation, and had not participated in Experiment 1 or 2.

Seven participants were excluded due to excessive artefacts. For the remaining 35 participants, average ERPs were computed over artefact-free trials for pronouns (match, M = 23, $S.D.$ = 4; mismatch, M = 23, $S.D.$ = 4), sentence-final words (match, M = 27, $S.D.$ = 2; mismatch, M = 27, $S.D.$ = 2).

Materials

Sixty sentences were selected from the set in the previous experiments. The subset was representative of the larger set in terms of male or female antecedent, proper name or noun phrase antecedent, high or low frequent antecedent. The sentences were extended with a further two words to make the sentences more similar in length to the new set of 160 filler sentences. The filler sentences were all grammatically correct, and semantically and referentially coherent. Manipulation-orthogonal comprehension questions were used to encourage participants to pay attention to the sentences. These questions appeared after 25%

of the sentences and were evenly divided across conditions.

Procedure

Participants were instructed to read for comprehension and answer the comprehension questions. The rest of the procedure (including Reading Span assessment), the data recording parameters and the statistical analysis were identical to that of Experiment 1 and 2. Due to the small number of pronoun sentences, however, no separate analyses were performed to test effects of antecedent type or anaphoric bias.

RESULTS

As shown in Figure 6 (upper pane), mismatching pronouns elicited a widely distributed, sustained negative shift started at about 200-300 ms and lasted throughout the whole epoch. There was no sign of the P600 effect that was observed in Experiment 1. At the sentence-final words (lower-left pane), the mismatch condition also did not elicit the N400 modulation that was seen in Experiment 1, but a positive deflection that started at about 600 ms. Of note, a separate results section that compared the ERP effects across experiments is available in the supplementary materials.

Pronoun results

300-600 ms: Mismatching pronouns elicited more negative ERPs across all electrodes (M = -1.08, $S.E.$ = .28, $F_{1,34}$ = 14.01, p = .001). The match effect was robust in all ROI subset analyses (all Fs > 11), but did not show any robust interaction with distribution.

600-1000: The main effects of match across all electrodes was marginally significant (M = -.91, $S.E.$ = .48, $F_{1,34}$ = 3.6, p = .07), where it was statistically significant in the medial analysis ($F_{1,34}$ = 4.5, p < .05) and the midline analysis ($F_{1,34}$ = 5.7, p < .05). In the crossline

analysis, there was a robust match by hemisphere effect was found ($F_{3,102} = 3.8$, $p < .05$): only a robust effect of match was observed at the central ROIs, LMC-ROI ($F_{1,34} = 5.4$, $p < .05$) and RMC ($F_{1,34} = 5.1$, $p < .05$).

1000-1500: The effect of match was significant across all electrodes ($M = -1.1$, $S.E. = .53$, $F_{1,34} = 4.2$, $p < .05$), marginally significant in the lateral analysis ($F_{1,34} = 3.0$, $p < .1$), and significant in the medial analysis ($F_{1,34} = 4.3$, $p < .05$), the midline analysis ($F_{1,34} = 6.7$, $p < .05$) and the crossline analysis ($F_{1,34} = 4.2$, $p < .05$), but no interactions with distributional factors were significant.

Reading Span results

Reading span scores ranged from 51 to 91 ($M = 70$, $S.D. = 10$). There was no significant correlation between reading span score and the difference between matching and mismatching pronouns LMFC-ROI in the 600-1000 time window (as was tested in the previous experiments).

Median split analyses showed no significant group effect ($F_{1,33} = .16$, *n.s.*), suggesting that the Nref effect was similarly strong in low and high span individuals.

Sentence-final word data

300-500 ms: The lateral analysis revealed a gender-match by anteriority by hemisphere 3-way interaction effect ($F_{3,102} = 4.8$, $p < .05$): in the LLFC-ROI the mismatch condition elicited more positive ERP responses than the match condition ($M = .60$, $S.E. = .27$, $F_{1,34} = 4.7$, $p < .05$), but not in any of the other ROIs. The subsequent time window was extended from 500-800 to 500-1000 ms to include the positive deflection that appeared nearer the end of the epoch.

500-1000 ms: The lateral analysis revealed a gender-match by anteriority by hemisphere 3-way interaction effect ($F_{3,102} = 3.44$, $p < .05$): the mismatch condition elicited more positive ERP responses than the match condition in the LLFC-ROI ($M = .97$, $S.E. = .35$, $F_{1,34} = 7.9$, $p < .01$), LLCP-ROI ($M = .79$, $S.E. = .29$, $F_{1,34} = 7.0$, $p < .05$), and the RAF-ROI ($M = .82$, $S.E. = .37$, $F_{1,34} = 5.0$, $p < .05$). The medial analysis revealed only a marginally significant effect of match ($F_{1,34} = 3.38$, $p < .1$). It should be noted, however, that the observed differences may have been due to pre-CW differences, as suggested by the very early divergence of the ERP waveforms between 0 and 100 ms after pronoun onset. If the Nref effect at the pronoun extended all the way to the sentence-final word, then possibly this sentence-final word effect is an artefact of the baseline procedure at this time point. Consistent with this idea, a novel baseline based on the 0-200 ms time window caused all the effects in the 300-500 and 500-1000 ms time window to disappear.

DISCUSSION

Mismatching pronouns in Experiment 3 also elicited a sustained negativity compared to matching pronouns. This negativity was much more widely distributed than in the Experiment 1 and 2. Nevertheless, the obtained effect was similar to the previous Nref effects in terms of its sustained nature as well as having a frontal maximum (even if only numerically in the earlier time windows). These results are a replication of Experiment 2, and suggest that a robust Nref is elicited even when mismatching pronoun are relatively infrequent in the experiment.

In contrast to Experiment 1 and 2, no relationship between the observed Nref effect and individual reading span score was observed. Individuals with lower or higher reading span score were equally likely to show an Nref effect. Only speculation can be offered regarding the lack of a group effect. It is possible that a significant correlation was harder to

detect due to larger variance in the observed ERP waveforms with fewer trials. Alternatively, it could be that the association between Nref and reading span simply is not robust enough.

The sentence-final effects in this experiment were qualitatively different from those in Experiment 2. A straightforward interpretation of these sentence-final effects and the comparison to sentence-final effects in Experiment 1 remains elusive, also because the sentences had been lengthened for Experiment 3, but they nevertheless appear to indicate that the effects of mismatch continue to impact comprehension downstream. As a caveat, however, it is possible that the observed effects stemmed from the baseline procedure. If the Nref effect extended downstream to the final word, which is indeed suggested by the waveforms, then ERP differences may have existed before the baseline period of the epoch used to analyse the sentence-final word effects. Consistent with this interpretation, an alternative baseline procedure that counteracted potential early differences (0-200 ms baseline procedure) in the ERP waveform made the later effects between 500 and 1000 ms disappear.

GENERAL DISCUSSION

Three ERP experiments examined the processing consequences of gender-mismatching pronouns (e.g., “The aunt found out that he had won the lottery”). Previous research has shown that these pronouns elicit P600 effects in sentences that are judged as unacceptable (Osterhout & Mobley, 1995). In each of the current experiments, mismatching pronouns elicited a sustained, frontal negative shift (Nref) compared to matching pronouns: when participants were instructed to hypothesize a new referent for mismatching pronouns (Experiment 1), and without this instruction (Experiment 2 and 3). In Experiment 1 and 2, this Nref effect was larger for individuals with high reading span, whereas both high and low span participants showed the effect in Experiment 3. Only in Experiment 1 was an additional

P600 effect of gender mismatch observed, which visually resembled P600 effects that were previously reported for pronoun gender mismatch (Nieuwland & Van Berkum, 2006; Osterhout & Mobley, 1995) and the P600 effects for syntactic violation filler sentences in Experiment 1 and 2. The pronoun mismatch P600 effect was robust only in the low reading span group, and when people strongly expected a matching pronoun. Moreover, current results replicated the association between pronoun-elicited P600 effects and sentence-final N400 effects reported by Osterhout and Mobley. The pronoun-elicited P600 effect-size in Experiment 1 predicted sentence-final N400 effect-size. In contrast to Osterhout and Mobley, the current study also performed exploratory analyses of pronoun mismatch as a function of antecedent type. In Experiment 1, noun phrase antecedents were associated with larger Nref effects, whereas proper name antecedents were associated with larger Nref effects in Experiment 2, although none of these exploratory analyses generated robust mismatch by antecedent type interaction effects.

Without an acceptability task that was used in previous work (Osterhout & Mobley, 1995), mismatching pronouns thus elicit qualitatively different ERP responses. The ambiguity inherently associated with an unknown referent gives rise to an Nref effect (Nieuwland & Van Berkum, 2008a, for review). The results thus suggest that people are more likely to interpret gender-mismatching pronouns as referring to an unknown antecedent rather than as a grammatically anomalous anaphor for a given antecedent, even without explicit instruction to do so. In other words, most participants may have attempted a 'referential solution' (i.e., assume a novel referent) instead of a 'syntactic solution' (i.e., assume that the pronoun had incorrect gender). Another important implication of this work is that whether people opt for the referential or syntactic solution depends on various factors, including individual differences in reading span, sentence meaning, experimental task instructions and possibly antecedent type. Arguably, the effects observed on ERP components in Experiments

2 and 3 imply processing biases without reliance on direct manipulation of task constraints via explicit instructions, as in Experiment 1 and in Osterhout and Mobley.

Unknown referents and the Nref

The most robust and main novel finding from the current series of experiments is that mismatching pronouns elicited a qualitatively distinct ERP effect compared to when participants explicitly judge such pronouns to be unacceptable (Osterhout & Mobley, 1995). The primary claim is thus that *under the various experimental conditions employed here* there is little evidence that people tend to perceive gender-mismatching pronouns as a syntactic anomaly (i.e., as an agreement error). The current results suggest that readers are more inclined to assign the antecedent retrieval difficulty to a problem at the level of the referential representation, rendering the pronoun ambiguous rather than anomalous.

These findings are consistent with discourse-based accounts of pronoun resolution (Garrod & Terras, 2000; Gerrig & O'Brien, 2005; Gordon & Hendrick, 1998; Kamp & Reyle, 1993; McKoon & Ratcliff, 1998; Myers & O'Brien, 1998; Sanford & Garrod, 1989). In these accounts, failure to retrieve an antecedent based on the pronoun's morphosyntactic features leads to additional retrieval processes (e.g., trying to find a suitable antecedent in the preceding discourse) or introduction of a novel discourse entity. In both cases, a referent must be assumed that is not already given in the sentence, but the current work does not directly address how elaborate or specific those new referent representations might be. The Nref effect may index only the initial perception of referential ambiguity from not knowing who the referent was. Alternatively, the Nref perhaps reflected an ongoing attempt to overcome this perceived ambiguity, perhaps by inferring the intended referent. Further research is needed to see if these alternative explanations of the Nref effect can be teased apart (see also Nieuwland & Berkum, 2008a, for discussion). However, the latter explanation is more

consistent with the sustained nature of the Nref and with previous Nref findings (e.g., Nieuwland & Van Berkum, 2008b), and is also suggested by the current findings.

It is important to point out that Nref effects were neither observed for all participants nor for all items. In Experiment 1, low reading span participants showed a P600 effect instead (see next section for discussion). In Experiment 2, however, low span participants showed no differential effect of pronoun mismatch, and the Nref effect hinged on high span participants. Notably, high span readers are more likely to elaborate their discourse models with optional, knowledge-based inferences than low span readers, whereas low span readers are more likely to entertain shallow representations of the discourse and are less inclined to resolve ambiguity (e.g., Linderholm, 2002; St. George et al., 1997; Whitney, Ritchie & Clark, 1991). The fact that only high span individuals consistently showed Nref effects suggests that the Nref may reflect a more effortful attempt to establish a referentially coherent interpretation. Under this interpretation, low span readers in Experiment 2 may have noticed gender mismatch, but neither attempted to infer the intended referent nor assumed that the pronoun had the wrong gender. The alternative interpretation that low span participants did not notice pronoun gender mismatch at all seems implausible, and inconsistent with the results from Experiment 1 and 3 where they showed robust effects of mismatch (albeit these showed different qualitative effects under different task constraints/instructions). It is unclear why Experiment 2 failed to show an effect. It may be that due to the lack of specific instructions and the large number of pronoun sentences, these participants failed to engage enough with the materials in a consistent way from trial to trial in Experiment 2. Consistent with this interpretation is the fact that low reading span participants in this experiment also showed smaller N400 effects for the filler semantic anomalies (see Supplementary Materials).

If low span readers in Experiment 2 failed to show a mismatch effect due to lack of elaborative processing, one can ask whether this explanation could apply to the lack of effect for noun phrase sentences in Experiment 2. It is possible that noun phrase antecedents caused participants to build less elaborate referential representations than the more specific proper names, because role descriptors receive less focus (Sanford et al., 1988). This is a speculative conclusion and requires support from further experimentation. An alternative explanation, in which the lack of Nref effects for noun phrase antecedents reflects easy introduction of a new referent, is incompatible with the larger Nrefs for high reading span participants. Importantly, the lack of robust match by antecedent type interaction effects precludes a firm conclusion, and the inconsistency of antecedent type effects experiments 1 and 2 could suggest that these results indicate type 1 error. A dedicated follow-up experiment is warranted, using sufficient trial numbers per antecedent type, while controlling for other potentially relevant influences such as anaphoric bias, antecedent frequency, as well as the extent to which noun phrase antecedents entail separate referents or not (e.g., 'mother' entails a child as it describes a relation between two people, whereas 'bachelor' does not).

Consistent with previous studies, the Nref effects emerged well before the onset of the next word and lasted throughout the epoch that was used for analysis. At the sentence-final word, however, the effect did not reappear. One reason for this could be the baseline procedure. If the Nref effect lasted until the end of the sentences but did not increase in size at the sentence-final word, then this effect will not appear in the sentence-final analysis. This is not to be taken as evidence that the referential processes indexed by the Nref are completed by the time of the sentence-final word, although this remains a possibility. The fact that no sentence-final effect was found for high reading span participants in Experiment 1 and all participants in Experiment 2 suggests that no additional processing costs are incurred at that point. The occurrence of a sentence-final effect in Experiment 3 seems inconsistent with this

conclusion, although this analysis may have suffered from a baseline problem. If indeed a robust sentence-final effect was elicited, however, it was a qualitatively distinct effect from the effect seen following a pronoun-elicited P600 effect.

Previous ERP work also suggests that the Nref reflects elaborative anaphoric inferences to establish referential coherence. For example, Nref effects occur only when two antecedents are equally plausible, not when two antecedents occur in the preceding discourse but one is more plausible than the other (e.g., Nieuwland et al., 2007; Nieuwland & Van Berkum, 2006). The current results suggest that people are more likely to engage in an anaphoric inference than to treat the pronouns as incorrect, at least when no acceptability judgment is required.

Below I consider the P600 effect in Experiment 1 and lack thereof in Experiment 2 and 3.

Pronoun mismatch and P600 effects

The explanation that I offer for the pronoun-elicited P600 effect in Experiment 1 adheres to the traditional association between syntactic processing difficulty and P600 modulations, which is based on the observation that syntactically unexpected or syntactically anomalous utterances reliably elicit P600 effects (e.g., Osterhout, McLaughlin, & Bersick, 1997b). Other extant literature, however, offers other interpretations of P600 effects, and most relevant to the current study is the claim that the P600 reflects the addition of a new referent (novel noun phrases compared to given noun phrases; Burkhardt, 2006, 2007; Schumacher & Hung, 2012; see also Kaan et al., 2007). Only Experiment 1 involved an explicit instruction to add new referents for mismatching pronouns, and only this experiment generated a P600 effect. Although I cannot speak to the extent to which participants followed task instructions in Experiment 1, which explicitly pointed out the syntactic mismatch

between pronoun and antecedent (see Discussion section of Experiment 1), these issues nevertheless beg the question whether the P600 effect reflected the addition of a new referent.

Importantly, the observed P600 effect has waveform and timing characteristics that are identical to those of P600 effects for straightforward syntactic anomalies (including the filler anomalies in this study), and, moreover, for previous pronoun gender-mismatch P600 effects (e.g., Nieuwland & Van Berkum, 2006; Osterhout & Mobley, 1995). In contrast, P600 effects that may reflect the adding of referents have different waveform morphology and timing characteristics (more peaked and shorted duration, e.g., Schumacher & Hung, 2012) than the effect reported here. Adding to that, the claim that low reading span individuals showed more signs of syntactic processing difficulty (i.e., larger P600 effects), at least under the particular task instruction in Experiment 1, is consistent with their tendency to opt for the most readily available interpretation (e.g., Linderholm, 2002). In single sentences with one antecedent, the readily available interpretation is a coreferential one. In contrast, the alternative interpretation that low span readers were more likely to add new referents than high span readers, is inconsistent with a wealth of data. Low reading span is associated with lower scores on off-line and on-line measures of language comprehension (e.g., Just and Carpenter, 1992) and the creation of less elaborate discourse models (e.g., Linderholm, 2002; St. George et al., 1997; Whitney et al., 1991). Furthermore, a new-referent interpretation of the P600 effect does not readily explain why the effect depended on the expectation of a matching pronoun, with somewhat larger effects when sentences biased readers towards a coreferential interpretation. Previous work shows P600 effects for pronouns that mismatched two given antecedents (Nieuwland & Van Berkum, 2006) or that mismatch the gender of the foregrounded antecedent, even with a matching non-foregrounded antecedent available (Van Berkum et al., 2007). Those effects are unlikely to reflect the addition of a novel referent,

whereas they are consistent with an interpretation that pronoun gender is initially taken to be incorrect.

Also problematic is the fact that the new-referent interpretation does not readily explain the association between the pronoun-elicited P600 effect and the sentence-final N400 effect. The observed sentence-final mismatch N400 effect replicated the Osterhout & Mobley findings, where only 'unacceptable'-responders showed both the pronoun-mismatch P600 effect and a sentence-final N400 effect. In the current study, larger pronoun-elicited P600 effects were significantly correlated with larger subsequent sentence-final N400 effects. Following the interpretation by Osterhout and Mobley (see also Osterhout et al., 1997a), sentence-final N400 effects could mean that by the end of the sentence a grammatically coherent analysis has not been reached. If the P600 effect reflected addition of a novel referent, it is unclear why a sentence-final N400 effect would follow.

Results from functional magnetic resonance imaging also fail to support the hypothesis that pronoun-elicited P600 effects reflect the adding of a referent (Nieuwland, Petersson & Van Berkum, 2007). Mismatching pronouns that elicit P600 ERP effects activate parietal brain regions that are sensitive to various morphosyntactic violations (e.g., Kuperberg et al., 2003; Nieuwland, Martin & Carreiras, 2012). In contrast, the neural generators for adding of referents lie in prefrontal brain regions associated with inference generation (e.g., Kuperberg, Lakshaman, Caplan & Holcomb, 2006).

A final point to make is that the association between P600 effects syntactic processing difficulty is well established (e.g., Osterhout et al., 1997b), while evidence for the P600 as index of the introduction of a new referent is inconsistent. Novel referents that are introduced via noun phrases *can* elicit a positive deflection when compared to given referents (Burkhardt, 2006, 2007). However, this effect does not appear consistently across similar

manipulations (e.g., Burkhardt & Roehm, 2007). ERP studies that compared new names to given names also did not find late positive ERP effects (e.g., Camblin, Ledoux, Boudewyn, Gordon & Swaab, 2007; Ledoux, Camblin, Gordon & Swaab, 2007). Moreover, other late positive ERP effects claimed to reflect the addition of referents differ greatly in timing and waveform morphology (e.g., Kaan et al.) and are perhaps more straightforwardly ascribed to syntactic garden-path phenomena. Hence, it remains to be seen whether late positive ERPs elicited by novel noun phrases actually reflect the addition of a new referent or instead reflect another aspect of comparing new information to given information.

For the reasons above, an explanation of the current P600 effect in terms of syntactic processing difficulty accounts for the results more parsimoniously and straightforwardly than a new-referent explanation. This is not to say that late positive deflections cannot reflect the addition of a new referent in other circumstances (e.g., introduction of referents via novel noun phrases). Importantly, different instantiations of late positive ERP effects exist in the psycholinguistic literature, which need not all reflect the same underlying processes (e.g., Bornkessel-Schlesewsky et al., 2008).

A final remark addresses the possibility that the observed P600 effects reflect a task-related decision. Previous research suggests that explicit judgement tasks elicit late positive ERP components (e.g., Kolk, Chwilla, Van Herten & Oor, 2003; Kuperberg, 2007; Roehm, Bornkessel-Schlesewsky, Rösler & Schlesewsky, 2007). It cannot be logically excluded from the current data that some or all of the observed P600 effect reflects the decision to add a referent. However, many of the arguments against the new-referent explanation above also hold against a decision-related explanation. For example, the results would then imply that low span participants performed the task whereas high reading span participants did not, which is implausible with respect to the literature on reading span and explicit measures of

language comprehension (e.g., Just & Carpenter, 1992). Reversely, if the Nref reflects the perceived ambiguity due to an extra-sentential referent and the P600 reflects an attempt at coreference, then high span participants indeed performed better on the task than the low span participants.

The conclusion that the observed P600 effect reflects the syntactic processing difficulty due to gender mismatch appears to be the most parsimonious explanation of the data. It suggests that low reading span participants attempted a coreferential interpretation, at least initially. Given the lack of a behavioral measure of task-performance, it remains an open question whether and when those participants indeed introduced a new referent.

The experimental circumstances in the current study are rather artificial, with a large number of unrelated and highly similar sentences being read from a computer screen. It is thus also an open question whether mismatching pronouns in regular conversation also give rise to a new referent being introduced or whether they would be treated as a syntactic anomaly. Comprehenders might by default assume that no error was made and that the speaker said what he/she intended (e.g. Grice, 1975; see also Nappa & Arnold, 2014). This makes sense, given that gender agreement errors that have consequences for the sex of the referent (such as pronoun gender in the current study) occur less often than gender agreement errors without consequences for the sex of the referent (Vigliocco & Frank, 1999). Because the observed effects depend on participant reading span, sentence meaning and task instructions, it stands to argue that there will be an effect of speaker identity too (e.g., Hanulíková, van Alphen, van Goch & Weber, 2012). The prediction is that if comprehenders have a reason to doubt the speaker's syntactic use or knowledge (e.g., when the speaker is a child or bilingual whose first language lacks gender agreement), they are more inclined to treat gender mismatch as a mistake rather than assume a different referent.

Conclusion

How people retrieve the antecedent for a referring expression has been, and probably still is one of the most elusive research topics in psycholinguistics (e.g., Clifton & Duffy, 2001; Garnham, 2001; Sanford & Garrod, 1989, for reviews). Pronouns have received special attention because they are purely referential devices that are inherently ambiguous. However, most of the sentence processing literature has ignored the fact that pronouns, besides referring back to given antecedents, can also introduce a new referent into the discourse representation (see Gerrig et al., 2011; Kitzinger et al., 2012, for corpus analysis; and for cataphora, see Kazanina et al., 2007; Van Gompel & Liversedge, 2003). The current ERP experiments followed up on the work of Osterhout and Mobley (1995) by comparing pronouns that matched or mismatched the only available antecedent. In absence of the acceptability judgment task that Osterhout and Mobley employed, mismatching pronouns elicited a very different pattern. Mismatching pronouns in each experiment elicited an Nref effect, the effect associated with referential ambiguity (Nieuwland & Van Berkum, 2008a). These results suggest that, at least under these circumstances, people are more likely to interpret mismatching pronouns as referring to an unknown antecedent rather than as a grammatically anomalous anaphor for a given antecedent.

REFERENCES

Almor, A., Kempler, D., MacDonald, M. C., Andersen, E. S., & Tyler, L. K. (1999). Why do Alzheimer patients have difficulty with pronouns? Working memory, semantics, and reference in comprehension and production in Alzheimer's disease. *Brain and Language*, *67*(3), 202-227.

Almor, A., & Nair, V. A. (2007). The form of referential expressions in discourse. *Language and Linguistics Compass*, *1*(1-2), 84-99.

Ariel, M. (1988). Referring and accessibility. *Journal of linguistics*, *24*(1), 65-87.

Arnold, J. E. (2010). How speakers refer: the role of accessibility. *Language and Linguistics Compass*, *4*(4), 187-203.

Arnold, J.E. (submitted). Women and men have different biases for pronoun interpretation.

Arnold, J. E., Eisenband, J. G., Brown Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course for pronoun resolution from eyetracking. *Cognition*, *76(1)*, B13-B26.

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 748.

Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, *46*(2), 391-418.

Benatar, A. & Clifton Jr, C. (2014). Newness, givenness and discourse updating: Evidence from eye movements. *Journal of Memory and Language*, 71(1), 1-16.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain research reviews*, *59*(1), 55-73.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, *98*(2), 159-168.

Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *Neuroreport*, *18*(17), 1851-1854.

Burkhardt, P., & Roehm, D. (2007). Differential effects of saliency: An event-related brain potential study. *Neuroscience letters*, *413*(2), 115-120.

Camblin, C. C., Ledoux, K., Boudewyn, M., Gordon, P. C., & Swaab, T. Y. (2007). Processing new and repeated names: Effects of coreference on repetition priming with speech and fast RSVP. *Brain Research*, *1146*, 172-184.

Chomsky, N. (1981). *Lectures on Government and Binding: The Pisa Lectures*. Holland: Foris Publications. 7th Edition. Berlin and New York: Mouton de Gruyter, 1993.

Clark, H. H., & Sengul, C. J. (1979). In search of referents for nouns and pronouns. *Memory & Cognition*, *7*(1), 35-41.

Clifton Jr, C., & Duffy, S. A. (2001). Sentence and text comprehension: Roles of linguistic structure. *Annual Review of Psychology*, *52*(1), 167-196.

Cunnings, I., Patterson, C., & Felser, C. (2014). Variable binding and coreference in sentence comprehension: Evidence from eye movements. *Journal of Memory and Language*, *71*(1), 39-56.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, *19*(4), 450-466.

Filik, R., Leuthold, H., Moxey, L. M., & Sanford, A. J. (2011). Anaphoric reference to quantified antecedents: An event-related brain potential study. *Neuropsychologia*, *49*(13), 3786-3794.

Filik, R., & Sanford, A. J. (2008). When is cataphoric reference recognised? *Cognition*, *107*(3), 1112-1121.

Filik, R., Sanford, A. J., & Leuthold, H. (2008). Processing pronouns without antecedents: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *20*(7), 1315-1326.

Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language*, *56*(3), 357-383.

Fukumura, K., Hyönä, J., & Scholfield, M. (2013). Gender Affects Semantic Competition: The Effect of Gender in a Non-Gender-Marking Language. *Journal of Experiment Psychology: Learning, Memory and Cognition*, 39(4), 1012-1021.

Fukumura, K., & van Gompel, R. P. (2010). Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language*, *62*(1), 52-66.

Fukumura, K., van Gompel, R. P., Harley, T., & Pickering, M. J. (2011). How does similarity-based interference affect the choice of referring expression? *Journal of Memory and Language*, *65*(3), 331-344.

Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Psychology Press.Garrod, S., Freudenthal, D., & Boyle, E. (1994). The role of different types of anaphor in the on-line resolution of sentences in a discourse. *Journal of memory and language*, *33*(1), 39-68.

Garrod, S., O'Brien, E. J., Morris, R. K., & Rayner, K. (1990). Elaborative inferencing as an active or passive process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 250.

Garrod, S.C., & Sanford, A.J. (1994). Resolving sentences in a discourse context: How discourse representation affects language understanding. In Gernsbacher (Ed), *Handbook of psycholinguistics. (pp. 675 698).* San Diego, CA, US: Academic Press.

Garrod, S., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. *Journal of Memory and Language*, *42*(4), 526-544.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, *32*(2), 99-156.

Gernsbacher, M. A., & Jescheniak, J. D. (1995). Cataphoric devices in spoken discourse. *Cognitive psychology*, *29*(1), 24-58.

Gerrig, R. J. (2005). The scope of memory-based processing. *Discourse Processes*, *39*(2-3), 225-242.

Gerrig, R. J., Horton, W. S., & Stent, A. (2011). Production and comprehension of unheralded pronouns: A corpus analysis. *Discourse Processes*, *48*(3), 161-182.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1-76.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive science*, *17*(3), 311-347.

Gordon, P. C., & Hendrick, R. (1998). The representation and processing of coreference in discourse. *Cognitive Science*, *22*(4), 389-424.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1411.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological review*, *101*(3), 371.

Greene, S. B., Gerrig, R. J., McKoon, G., & Ratcliff, R. (1994). Unheralded pronouns and management by common ground. *Journal of Memory and Language*, *33*(4), 511-526.

Greene, S. B., McKoon, G., & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 266.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), Syntax and semantics 3: Speech acts (pp. 41–58). New York: Academic Press.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 274-307.

Hammer, A., Jansma, B. M., Lamers, M., & Münte, T. F. (2008). Interplay of meaning, syntax and working memory during pronoun resolution investigated by ERPs. *Brain research*, *1230*, 177-191.

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*(4), 878-887.

Harris, T., Wexler, K., & Holcomb, P. (2000). An ERP investigation of binding and coreference. *Brain and Language*, *75*(3), 313-346.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological review*, *99*, 122-149.

Kaan, E., Dallas, A. C., & Barkley, C. M. (2007). Processing bare quantifiers in discourse. *Brain research*, *1146*, 199-209.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and cognitive processes*, *15*(2), 159-201.

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, *15*(1), 98-110.

Kamp, H., & Reyle, U. (1993). *From discourse to logic: Introduction to model theoretic semantics of natural language, formal logic and discourse representation theory*. Springer.

Kazanina, N., Lau, E. F., Lieberman, M., Yoshida, M., & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language*, *56*(3), 384-409.

Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics*, *39*(1-2), 1-37.

King, J. W., & Kutas, M. (1997). Is she an engineer? Brain potentials and anaphora. In *Poster presented at the Fourth Annual Meeting of the Cognitive Neuroscience Society, Boston*.

Kitzinger, C., Shaw, R., & Toerien, M. (2012). Referring to persons without using a full-form reference: Locally initial indexicals in action. *Research on Language & Social Interaction*, *45*(2), 116-136.

Kolk, H. H., Chwilla, D. J., van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and language*, *85*(1), 1-36.

Koornneef, A. W., & Sanders, T. J. (2012). Establishing coherence relations in discourse: The influence of implicit causality and connectives on pronoun resolution. *Language and Cognitive Processes*, (ahead-of-print), 1-38.

Koornneef, A. W., & Van Berkum, J. J. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, *54*(4), 445-465.

Kreiner, H., Sturt, P., & Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language*, *58*(2), 239-261.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23-49.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience*, *15*(2), 272-293.

Kuperberg, G. R., Lakshmanan, B. M., Caplan, D. N., & Holcomb, P. J. (2006). Making sense of discourse: An fMRI study of causal inferencing across sentences. *Neuroimage*, *33*(1), 343-361.

Kutas, M., Federmeier, K. D., Coulson, S., King, J. W., Munte, T. F. Language, In: J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson (Eds.), *Handbook of Psychophysiology*, Cambridge University Press, 2000, pp. 576-601.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203-205.

Lamers, M. J., Jansma, B. M., Hammer, A., & Münte, T. F. (2006). Neural correlates of semantic and syntactic processes in the comprehension of case marked pronouns: evidence from German and Dutch. *BMC neuroscience*, *7*(1), 23.

Ledoux, K., Gordon, P. C., Camblin, C. C., & Swaab, T. Y. (2007). Coreference and lexical repetition: Mechanisms of discourse integration. *Memory & cognition*, *35*(4), 801-815.

Levine, W. H., Guzmán, A. E., & Klin, C. M. (2000). When anaphor resolution fails. *Journal of Memory and Language*, *43*(4), 594-617.

Love, J., & McKoon, G. (2011). Rules of engagement: Incomplete and complete pronoun resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 874.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109(1), 35-54.

Martin, A. E., & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language*, *58*(3), 879-906.

Martin, A. E., & McElree, B. (2011). Direct-access retrieval during sentence comprehension: Evidence from Sluicing. *Journal of memory and language*, *64*(4), 327-343.

Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2012). Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *Neuroimage*, *59*(2), 1859-1869.

Martin, A.E., Nieuwland, M.S., & Carreiras, M. (2014). Agreement attraction during comprehension of grammatical sentences: ERP evidence from ellipsis. *Brain and Language.* DOI: *10.1016/j.bandl.2014.05.001*

McDonald, J. L., & Shaibe, D. M. (2002). The accessibility of characters in single sentences: Proper names, common nouns, and first mention. *Psychonomic bulletin & review*, *9*(2), 356-361.

McKoon, G., & Ratcliff, R. (1998). Memory-based language processing: Psycholinguistic research in the 1990s. *Annual Review of Psychology*, *49*(1), 25-42.

Murphy, G. L. (1984). Establishing and accessing referents in discourse. *Memory & Cognition*, *12*(5), 489-497.

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, *26*(2-3), 131-157.

Nappa, R., & Arnold, J. E. (2014). The road to understanding is paved with the speaker's intentions: Cues to the speaker's attention and intentions affect pronoun comprehension. *Cognitive psychology*, *70*, 58-81.

Nieuwland, M. S., Martin, A. E., & Carreiras, M. (2012). Brain regions that process case: evidence from Basque. *Human brain mapping*, *33*(11), 2509-2520.

Nieuwland, M. S., Otten, M., & Van Berkum, J. J. (2007). Who are you talking about? Tracking discourse-level referential processing with event-related brain potentials. *Journal of Cognitive Neuroscience*, *19*(2), 228-236.

Nieuwland, M. S., Petersson, K. M., & Van Berkum, J. J. (2007). On sense and reference: Examining the functional neuroanatomy of referential processing. *Neuroimage*, *37*(3), 993-1004.

Nieuwland, M. S., & Van Berkum, J. J. (2006). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research*, *1118*(1), 155-167.

Nieuwland, M. S., & Van Berkum, J. J. (2008a). The neurocognition of referential ambiguity in language comprehension. *Language and Linguistics Compass*, *2*(4), 603-630.

Nieuwland, M. S., & Van Berkum, J. J. (2008b). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from ERPs. *Brain and language*, *106*(2), 119-131.

Osterhout, L. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes*, *14*(1), 1-14.

Osterhout, L., Bersick, M., & McLaughlin, J. (1997a). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, *25*(3), 273-285.

Osterhout, L., McLaughlin, J., & Bersick, M. (1997b). Event-related brain potentials and human language. *Trends in Cognitive Sciences*, *1*(6), 203-209.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785-806.

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 786.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and language, 34*(6), 739-773.

Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes*, *14*(3), 283-317.

Qiu, L., Swaab, T. Y., Chen, H. C., & Wang, S. (2012). The role of gender information in pronoun resolution: Evidence from Chinese. *PloS one*, *7*(5), e36156.

Rohde, H. & Kehler, A. (in press). Grammatical and information-structural influences on pronoun production. *Language and Cognitive Processes: Special Issue on Production of Referring Expressions.*

Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., & Schlesewsky, M. (2007). To predict or not to predict: Influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience*, *19*(8), 1259-1274.

Sanford, A. J., & Garrod, S. C. (1989). What, when, and how?: Questions of immediacy in anaphoric reference resolution. *Language and Cognitive Processes*, *4*(3-4), SI235-SI262.

Sanford, A. J., Garrod, S., Lucas, A., & Henderson, R. (1983). Pronouns without explicit antecedents? *Journal of Semantics*, *2*(3-4), 303-318.

Sanford, A. J., Moar, K., & Garrod, S. C. (1988). Proper names as controllers of discourse focus. *Language and speech*, *31*(1), 43-56.

Schumacher, P. B., & Hung, Y. C. (2012). Positional influences on information packaging: Insights from topological fields in German. *Journal of Memory and Language*, *67*(2), 295-310.

St. George, M., Mannes, S., & Hoffman, J. E. (1997). Individual differences in inference generation: An ERP analysis. *Journal of Cognitive Neuroscience*, *9*(6), 776-787.

Streb, J., Rösler, F., & Hennighausen, E. (1999). Event-related responses to pronoun and proper name anaphors in parallel and nonparallel discourse structures. *Brain and Language*, *70*(2), 273-286.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*(3), 542-562.

Sturt, P. (2013). Referential processing in sentences. In Gompel, R. V. (Ed.), *Sentence Processing.* (1 ed.) (Current issues in the Psychology of Language). Psychology Press.

Swaab, T. Y., Camblin, C. C., & Gordon, P. C. (2004). Electrophysiological evidence for reversed lexical repetition effects in language processing. *Journal of Cognitive Neuroscience*, *16*(5), 715-726.

Van Berkum, J. J., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of memory and language*, *41*(2), 147-182.

Van Berkum, J. J., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research*, *1146*, 158-171.

van den Noort, M., Bosch, P., Haverkort, M., & Hugdahl, K. (2008). A standard computerized version of the Reading Span Test in different languages. *European Journal of Psychological Assessment*, *24*(1), 35.

van Gompel, R. P., & Liversedge, S. P. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(1), 128.

van Rij, J., van Rijn, H., & Hendriks, P. (2011). WM load influences the interpretation of referring expressions. In *Proceedings of the 2nd workshop on Cognitive Modeling and Computational Linguistics* (pp. 67-75). Association for Computational Linguistics.

Vigliocco, G., & Franck, J. (1999). When sex and syntax go hand in hand: Gender agreement in language production. *Journal of Memory and Language*, *40*(4), 455-478.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, *85*(1), 79-112.

Whitney, P., Ritchie, B. G., & Clark, M. B. (1991). Working-memory capacity and the use of elaborative inferences in text comprehension. *Discourse Processes*, *14*(2), 133-145.

Xu, X., Jiang, X., & Zhou, X. (2013). Processing biological gender and number information during Chinese pronoun resolution: ERP evidence for functional differentiation. *Brain and cognition*, *81*(2), 223-236.

Yekovich, F. R., & Walker, C. H. (1978). Identifying and using referents in sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, *17*(3), 265-277.

FIGURE CAPTIONS

FIGURE 1. Electrode configuration (black letters) and the Region-of-Interest clusters that were used for statistical analysis (white letters). The last one or two letters refer to the anterior/posterior dimension: AF = Anterior Frontal, FC = Frontocentral, C = Central, CP = Centroparietal, PO = Parieto-occipital. The first letter of 3-letter cluster-names and the first two letters of 4-letter cluster names refer to left-right dimension: L/R = Left/Right, LL/RL = Left/Right Lateral, LM/RM = Left/Right Medial.

FIGURE 2. Results from Experiment 1, upper pane: Pronoun gender-match effects at 10 selected channels, and the scalp distribution of the mismatch minus match differential effect in the time windows used for statistical analysis. Lower left pane: ERPs at Pz elicited by sentence-final words in match and mismatch conditions, and the scalp distributions in the N400 and P600 ms time window. Lower right pane: ERPs at Pz elicited by filler sentences and scalp distribution of the semantic anomaly and syntactic violation effect (each compared to the correct control sentences). All ERP waveforms are filtered at 5 Hz high cut-off for presentation purposes. Note that negativity is plotted upwards. Additional figures that show all electrodes are available as Supplementary Materials on the JML website.

FIGURE 3. Pronoun gender-match effects at electrode F3, P3 and PO3 for sentences with noun phrase antecedents or proper name antecedents in Experiment 1 and Experiment 2, along with the scalp distributions of the difference between match and mismatch conditions in the 300-600 ms time window.

FIGURE 4. Pronoun gender-match effects at electrode F3, P3 and PO3 for sentences with relatively high anaphoric bias or relatively low anaphoric bias in Experiment 1 and Experiment 2, along with the scalp distributions of the difference between match and mismatch conditions in the 600-1000 ms time window.

FIGURE 5. Results from Experiment 2, upper pane: Pronoun gender-match effects at selected channels, and the scalp distribution of the mismatch minus match differential effect in the time windows used for statistical analysis. Lower left pane: ERPs at Pz elicited by sentence-final words in match and mismatch conditions, and the scalp distributions in the N400 and P600 ms time window. Lower right pane: ERPs at Pz elicited by filler sentences and scalp distribution of the semantic anomaly and syntactic violation effect (each compared to the correct control sentences).

FIGURE 6. Results from Experiment 3, upper pane: Pronoun gender-match effects at selected channels, and the scalp distribution of the mismatch minus match differential effect in the time windows used for statistical analysis. Lower left pane: ERPs at Pz elicited by sentence-final words in match and mismatch conditions, and the scalp distributions in the N400 and P600 ms time window.

ACKNOWLEDGEMENTS

I am grateful to Keelin Murray for help with the materials, to Chrysa Retsa, Aine Ito, Rachel King, Juhani Virta, Prerana Sabnis and Amy Cowie for help with data collection. I also wish to thank Chuck Clifton, Andrea Eyleen Martin-Nieuwland, Patrick Sturt and three anonymous reviewers for providing helpful comments on previous drafts of this manuscript. This work was funded by a PPLS Pilot Project Funds awarded by the School of Philosophy, Psychology and Language Sciences from the University of Edinburgh.

APPENDIX

All items from Experiment 1 and 2 are listed below (pronoun match conditions only), with anaphoric bias scores for the pronoun sentences. The second column give the antecedent condition (first character m/f for male/female, second and third character np/pn for noun phrase/proper name, fourth and fifth character hf/lf fir high frequent/low frequent).

1	mnphf	The boy thought that he would win the race.	4
2	mnphf	The boyfriend said that he was not being faithful.	12
3	mnphf	The brother claimed that he didn't steal the money.	4
4	mnphf	The butler knew that he would be busy later.	3
5	mnphf	The father hoped that he would feel better tomorrow.	-20
6	mnphf	The fireman denied that he had been in danger.	19
7	mnphf	The groom said that he regretted drinking last night.	19
8	mnphf	The guy realised that he might miss the bus.	16
9	mnphf	The husband insinuated that he knew about the affair.	3
10	mnphf	The king regretted that he had declared war again.	22
11	mnphf	The lad worried that he would not earn enough.	13
12	mnphf	The man noted that he was not always honest.	-5
13	mnphf	The nephew forgot that he had bought a takeaway.	-6
14	mnphf	The policeman insisted that he was in big trouble.	-3
15	mnphf	The prince predicted that he would lose the competition.	5
16	mnphf	The salesman assumed that he would get a bonus.	-6
17	mnphf	The schoolboy proved that he had done the homework.	18
18	mnphf	The son ordered that he should get new shoes.	0
19	mnphf	The uncle heard that he was not employed anymore.	-19
20	mnphf	The waiter believed that he was getting bad tips.	2
21	mpnhf	Adam noticed that he had no milk left.	2
22	mpnhf	Andy hinted that he would tell the truth.	20
23	mpnhf	Brian said that he wouldn't attend the wedding.	21
24	mpnhf	Daniel was surprised that he has heard the gossip.	-5
25	mpnhf	David promised that he would make it home.	23
26	mpnhf	Harry was relieved that he passed the maths test.	8
27	mpnhf	Harry revealed that he was not in love.	22
28	mpnhf	Jack announced that he felt sick after dinner.	24
29	mpnhf	James asked whether he could close the window.	8
30	mpnhf	John shouted that he was very angry today.	19
31	mpnhf	Kevin remarked that he didn't have any plans.	8
32	mpnhf	Mark responded that he had no money anyway.	17
33	mpnhf	Michael requested that he might cook some dinner.	12
34	mpnhf	Oliver imagines that he is popular with colleagues.	14
35	mpnhf	Richard perceived that he embarrassed the poor nurse.	-1
36	mpnhf	Robert saw that he had parked badly yesterday.	-7
37	mpnhf	Stephen recorded that he didn't understand the question.	3
38	mpnhf	Tim suggested that he should go home early.	-6
39	mpnhf	Tom concluded that he needed help writing essays.	6
40	mpnhf	William confirmed that he would need surgery immediately.	11
41	fnphf	The actress inferred that she would need more lessons.	19
42	fnphf	The aunt intimated that she was not happily married.	10
43	fnphf	The bride observed that she looked worn out today.	-10

44	fnphf	The daughter alleged that she had stolen the car.	4
45	fnphf	The girl affirmed that she understood the rules well.	10
46	fnphf	The girlfriend maintained that she hadn't done anything wrong.	1
47	fnphf	The housewife mentioned that she might go shopping later.	18
48	fnphf	The lady stated that she heard nothing last night.	13
49	fnphf	The maid asserted that she had bought the flowers.	16
50	fnphf	The mother noticed that she had forgotten my birthday.	-22
51	fnphf	The nanny indicated that she felt much better today.	-13
52	fnphf	The niece recalled that she had promised to drive.	1
53	fnphf	The princess was reminded that she was not old enough.	19
54	fnphf	The queen was assured that she was the best looking.	-3
55	fnphf	The saleswoman was warned that she was selling too little.	-8
56	fnphf	The schoolgirl whispered that she had a massive crush.	25
57	fnphf	The sister muttered that she was angry and sad.	0
58	fnphf	The waitress discovered that she served the wrong food.	4
59	fnphf	The wife wished that she could remember dates better.	-4
60	fnphf	The woman suspected that she might not get better.	-14
61	fpnhf	Amy supposed that she should rent a car.	3
62	fpnhf	Anne presumed that she would ace the exam.	4
63	fpnhf	Danielle expected that she could order pizza later.	-1
64	fpnhf	Emily anticipated that she might get in trouble.	8
65	fpnhf	Emma understood that she had won the lottery.	19
66	fpnhf	Jane mentioned that she would go collect Mum.	17
67	fpnhf	Jenny replied that she had gotten the job.	24
68	fpnhf	Linda estimated that she had spent ninety pounds.	6
69	fpnhf	Julia detected that she felt unsafe after dark.	-11
70	fpnhf	Kate identified that she was secretly the culprit.	-1
71	fpnhf	Laura foresaw that she would fall in love.	1
72	fpnhf	Lily established that she would not eat dessert.	2
73	fpnhf	Lisa complained that she had slipped on ice.	4
74	fpnhf	Mary remembered that she had dry-cleaning to collect.	17
75	fpnhf	Nicole vowed that she would get revenge soon.	23
76	fpnhf	Rachel reported that she was attacked at school.	1
77	fpnhf	Sarah was convinced that she was right to complain.	-3
78	fpnhf	Lucy reckoned that she should sleep in tomorrow.	10
79	fpnhf	Susan felt that she was in the wrong.	19
80	fpnhf	Victoria gathered that she must apologise straight away.	-2
81	mnplf	The alderman inferred that he might fall under suspicion.	12
82	mnplf	The bachelor intimated that he was looking for love.	16
83	mnplf	The businessman observed that he was very successful recently.	0
84	mnplf	The butcherboy alleged that he hadn't seen the crime.	3
85	mnplf	The choirboy affirmed that he had practised all weekend.	22
86	mnplf	The craftsman maintained that he used the best materials.	9
87	mnplf	The duke mentioned that he was purchasing another home.	18
88	mnplf	The Dutchman stated that he was far from home.	17
89	mnplf	The Englishman asserted that he was happy and proud.	14
90	mnplf	The fisherman pretended that he would deliver it tomorrow.	27
91	mnplf	The footman indicated that he knew the customer well.	7
92	mnplf	The foreman recalled that he had seen someone leave.	2
93	mnplf	The Frenchman was reminded that he had duties to perform.	15

94	mnplf	The gentleman was assured that he would get home safely.	11
95	mnplf	The horseman was warned that he might meet bad weather.	2
96	mnplf	The huntsman whispered that he could keep a secret.	14
97	mnplf	The landlord muttered that he needed money right now.	1
98	mnplf	The marksman discovered that he couldn't see very well.	10
99	mnplf	The spokesman wished that he didn't have to speak.	10
100	mnplf	The weatherman suspected that he might be wrong today.	1
101	mpnlf	Balthazar supposed that he would leave work soon.	5
102	mpnlf	Byron presumed that he would be happier elsewhere.	0
103	mpnlf	Casimir expected that he would be hung-over tomorrow.	8
104	mpnlf	Cassius anticipated that he would retire next year.	6
105	mpnlf	Claude understood that he was an excellent baker.	12
106	mpnlf	Clifford mentioned that he was getting a divorce.	20
107	mpnlf	Cormac replied that he could start a business.	18
108	mpnlf	Donatello estimated that he would lose the deposit.	2
109	mpnlf	Errol detected that he had been cruel yesterday.	-6
110	mpnlf	Juan identified that he was really the thief.	6
111	mpnlf	Godfrey foresaw that he would be acquitted immediately.	1
112	mpnlf	Heath established that he had a solid alibi.	16
113	mpnlf	Jeremiah complained that he was getting no attention.	3
114	mpnlf	Reynold remembered that he had promised to attend.	21
115	mpnlf	Romulus vowed that he would avenge this death.	23
116	mpnlf	Russell reported that he had a great story.	5
117	mpnlf	Vincenzo was convinced that he had won the debate.	3
118	mpnlf	Warren reckoned that he was very funny yesterday.	7
119	mpnlf	Xander felt that he was working too hard.	18
120	mpnlf	Xavier gathered that he was being gossiped about.	-3
121	fnplf	The ballerina thought that she had pulled a muscle.	15
122	fnplf	The bridesmaid said that she had enjoyed the wedding.	8
123	fnplf	The businesswoman claimed that she was earning too little.	17
124	fnplf	The chambermaid knew that she was running late today.	4
125	fnplf	The choirgirl hoped that she would pass the audition.	16
126	fnplf	The duchess denied that she was wasteful with money.	26
127	fnplf	The empress claimed that she had bought the yacht.	11
128	fnplf	The Englishwoman realised that she had sold no jam.	12
129	fnplf	The fisherwoman insinuated that she had lied about quotas.	-5
130	fnplf	The Frenchwoman regretted that she had been caught lying.	28
131	fnplf	The governess worried that she might not get paid.	-2
132	fnplf	The heiress noted that she should donate to charity.	3
133	fnplf	The landlady forgot that she had taken a deposit.	11
134	fnplf	The policewoman insisted that she had seen the crime.	-9
135	fnplf	The priestess predicted that she would pray all night.	-5
136	fnplf	The showgirl assumed that she would earn great tips.	9
137	fnplf	The spinster proved that she was very efficient actually.	13
138	fnplf	The spokeswoman ordered that she receive all the newspapers.	-11
139	fnplf	The washerwoman heard that she would get a raise.	-7
140	fnplf	The weathergirl believed that she had taken a risk.	4
141	fpnlf	Adriana noticed that she was putting on weight.	-3
142	fpnlf	Althea hinted that she wanted a new puppy.	15
143	fpnlf	Alyssa said that she should work out more.	22

144 fpnlf Cassandra was surprised that she enjoyed the terrible movie. 2
145 fpnlf Celeste promised that she was not a liar. 20
146 fpnlf Claudette was relieved that she had been rescued safely. 8
147 fpnlf Cordelia revealed that she had a secret past. 22
148 fpnlf Elizabella announced that she would not ever marry .27
149 fpnlf Evelina asked that she be excused from dinner. 6
150 fpnlf Giselle shouted that she had been lied to. 17
151 fpnlf Lavinia remarked that she had run very quickly. 1
152 fpnlf Layla responded that she had enjoyed the meal. 16
153 fpnlf Nerissa requested that she might take a nap. 7
154 fpnlf Octavia imagined that she was liked by everyone. 18
155 fpnlf Regina perceived that she had missed the joke. -5
156 fpnlf Sabrina saw that she had reached the peak. -4
157 fpnlf Tirza recorded that she was feeling much better. 3
158 fpnlf Ursula suggested that she might buy a laptop. 1
159 fpnlf Valentina concluded that she would have to leave. 5
160 fpnlf Roselle confirmed that she was learning to fly. 16

Filler sentences (correct/syntactic violation/semantic violation)

1 Children like to read/reading/manage funny cartoons.
2 The beavers sometimes chew/chewing/melt the garden hose.
3 Car crashes can delay/delaying/learn traffic for hours.
4 The peanuts will make/making/grill you feel full.
5 The powerful magnets will pull/pulling/learn defective parts.
6 The new textbook could draw/drawing/hear various students.
7 The new software will print/printing/glue very elaborate pictures.
8 The daily newspapers should land/landing/dance on the porch.
9 Those small spiders would often spin/spinning/burn beautiful webs.
10 The little dog always waits/waiting/peaks in the driveway.
11 A new computer will last/lasting/paint for many years.
12 The weather is set to improve/improving/whistle next week.
13 Manuela brought her book to read/reading/sniff all night.
14 When gardening, you must water/watering/glue the plants carefully.
15 My father's dancing might shock/shocking/stab you a bit.
16 This ointment will cure/curing/loathe all forms of skin disease.
17 The boulder seemed to rest/resting/live precariously on the mountain.
18 The simulated accident might frighten/frightening/ignore the children very much.
19 Simple vegetable oil is used to fry/frying/plow the vegetables.
20 The strawberry beds might tempt/tempting/sneeze rabbits and other animals.
21 Where the road forks/forking/believes is where I got lost.
22 Alison used a hammer to break/breaking/kiss the small lock.
23 This book will explain/explaining/dance the history of the monarchy.
24 To knit a scarf I start/starting/cry with finding textiles.
25 The cranky babies will have to sleep/sleeping/explode later on.
26 The dog always wants to walk/walking/discuss in the evenings.
27 The cats won't eat/eating/bake the food that Petra buys.
28 This exotic spice may add/adding/seek the flavour I enjoy.
29 The pacifier we bought will soothe/soothing/drop the cranky baby.
30 The security camera will now take/taking/trip photographs of everyone.

31 Betsy went out to pick/picking/melt apples for a pie.
32 The newly planted grass will grow/growing/swim quite a bit.
33 The alarm system should warn/warning/swear that there is an intruder.
34 Heather knew that the hotel food would cost/costing/fight too much.
35 This rare herb can heal/healing/count the pains in your back.
36 The fingerprints on the gun could prove/proving/judge the defendant's innocence.
37 The black widow spider likes to hide/hiding/sigh in dark places.
38 One kangaroo at the zoo would sometimes sit/sitting/write all day.
39 The award winning play will run/running/leap for several more months.
40 The hiker used his last match to start/starting/tie the fire.
41 To fix the tyres Dad had to pump/pumping/knit them carefully.
42 The doctor announced that the symptoms would last/lasting/knit until Christmas.
43 The boxes in the attic may still hold/holding/find many photographs.
44 Susan worried that her kitten would scratch/scratching/lift the young child.
45 I thought that I would fit/fitting/bark right in with them.
46 The chemical additive may tend/tending/desire to lower the freezing point.
47 The colours should not fade/fading/walk when the sweater is washed.
48 The many bugs must eat/eating/buy a head of lettuce hourly.
49 The therapist hoped that the drug would calm/calming/clean the anxious patient.
50 Critics say that rap songs tend/tending/learn to lead young people astray.
51 The assistant was told that the alibi would prevent/preventing/cook an indictment.
52 The booklet says that contraceptives will fail/failing/complain if used too sparingly.
53 The new romance novel should sell/selling/clean in every store this year.
54 The new crop of corn should feed/feeding/scrape everyone in the state.
55 With my new bike, I can cycle/cycling/crawl to work every day.
56 My new dogs will eat/eating/type all day if I let them.
57 The tree in the garden can’t sprout/sprouting/sell new buds in this weather.
58 The elephants get on their hind legs and stand/standing/chirp to impress audiences.
59 The account of the incident didn’t match/matching/paste the one given by others.
60 The general admits that the missile might explode/exploding/call before leaving the area.
61 Harold’s rubber raft will hit/hitting/gloves a rock.
62 This pen will not write/writing/sing any more.
63 Mother will have to cook/cooking/burp all day.
64 The chambermaid will clean/cleaning/hit your room today.
65 The baby’s pacifier might hurt/hurting/cheat too much.
66 This old electric blender doesn't mix/mixing/own smoothies anymore.
67 There are otters that swim/swimming/fly and do tricks.
68 Landfill chemicals may mix/mixing/hope to create lethal substances.
69 These grapevines don’t grow/growing/jog well in sandy regions.
70 The raging bull will charge/charging/whistle at the man.
71 The local beers will satisfy/satisfying/trip every beer drinker.
72 Scary things come out to play/playing/melt at Halloween.
73 New born lambs often frolic/frolicking/question around the fields.
74 The flowers will brighten/brightening/type up the hospital ward.
75 This test might fail/failing/hate to discriminate among students.
76 The puppy seems to like/liking/call sleeping during the day.
77 The new drugs can prevent/preventing/drink many forms of disease.
78 The new toothpaste could help/helping/beg to provide proper protection.
79 The portrait of Uncle Henry doesn’t look/looking/sing like him.

80 The bull that escaped could smash/smashing/dance the wooden fence.
81 The pet cats will soon eat/eating/describe their evening meal.
82 My office doesn't smell/smelling/paint since I opened the window.
83 The caged lion will roar/roaring/sing at tourists passing by.
84 I often wish everyone loved to dance/dancing/staple all night.
85 My new boss expects me to type/typing/chew all day.
86 My athlete sister will run/running/stab a marathon next month.
87 The astronomer might prove/proving/shout that the moon has canals.
88 The fighter plane can fly/flying/walk faster than anyone expects.
89 The red ants will bite/biting/wash if you aren't careful.
90 Trudy found it difficult to drive/driving/boil for several months.
91 The lever does not shut/shutting/lift off the power supply.
92 The falcon chicks always chirp/chirping/staple until they are fed.
93 The new species of orchid will grow/growing/sing in tropical regions.
94 The composer hoped his music would enchant/enchanting/question the public.
95 The farmhouse is so old that it scares/scaring/writes the neighbours.
96 The French clock doesn't tell/telling/ask the time during power failures.
97 The report mentioned that factories should train/training/hug workers more thoroughly.
98 The sea lions can bask/basking/edit on the beach all day.
99 The little baby sneezes/sneezing/types so much, it needs a doctor.
100 The noisy ducks will soon fly/flying/skip away from the lake.
101 The new television didn't look/looking/sneeze too difficult to set up.
102 My favourite pub is too busy to sit/sitting/own in now.
103 We hoped the award would cheer/cheering/wash up the depressed student.
104 The hidden door will open/opening/cook when the code is spoken.
105 The plumber said that water might seep/seeping/speak from the refrigerator.
106 My TV had better not break/breaking/trip before I watch House.
107 The teacher said our report must not last/lasting/cry very long.
108 Billy bumped his bicycle, causing it to fall/falling/sneeze into the street.
109 I think that the leaky tub might bother/bothering/ask the tenants downstairs.
110 These pens shouldn't be used to sketch/sketching/dust but only for writing.
111 People hope that the sculptures will inspire/inspiring/invent new forms of art.
112 The new heater in the maid's room should dry/drying/find the laundry.
113 The new detergent is supposed to clean/cleaning/burn the floors with ease.
114 Margaret was concerned that her cats might shed/shedding/cheat on her sofa.
115 Tea is a drink that people often drink/drinking/scrape in the afternoon.
116 The cowboy gives his horse a chance to drink/drinking/fish from the stream.
117 The skyscraper being built by the city would block/blocking/send out the sunlight.
118 It was hard to get the infant to smile/smiling/vote for the photographer.
119 The movers thought that the piano would weigh/weighing/cough less than it did.
120 My brother bet that this spider could climb/climbing/type faster than you could.

Additional semantically and syntactically unproblematic fillers

1 Europeans often dislike American tourists.
2 Many artists paint with watercolours.
3 Most textbooks have extensive indexes.
4 Many hurricanes start in the Caribbean.
5 Shopping centres have become increasingly popular.

6 Most butchers cut mean using cleavers.
7 Florida alligators like to eat raw hamburger.
8 Politicians take two month holidays every summer.
9 Most kittens claw the furniture for fun.
10 Hot liquids become gas at specific temperatures.
11 Polar bears live at the North Pole.
12 The ski slopes in Austria are very challenging.
13 My clothes tend to change with fashion trends.
14 Few lawyers donate their time to the poor
15 Most dentists recommend brushing your teeth twice daily.
16 Few gardeners know how to grown exotic flowers.
17 Car keys have a way of getting lost easily.
18 The media hopes that political campaigns are close contests.
19 The newly elected officials hope to balance the budget.
20 Board games have become less popular in recent years.
21 Baby chimpanzees make terrible pets.
22 Old elevators have creaky doors.
23 African elephants live in the jungle.
24 Marathons attract the best runners around.
25 Pacific Islands are excellent holiday destinations.
26 Trains are more comfortable than buses.
27 Movie directors make more money than actors.
28 Few students know how to study anymore.
29 Many memories fade after a few years.
30 Universities hope to recruit more foreign students.
31 I often buy fresh strawberries before work.
32 Young fathers want more time off from work.
33 Most juries agree on a verdict in hours.
34 Most meteors burn up before they reach earth.
35 Modern office buildings often resemble sheets of glass.
36 These flowers wilt when left above the radiator.
37 Many magicians know how to escape from a safe.
38 The reluctant witness seems to be holding up well.
39 Political candidates travel all over the country on campaign.
40 The car would start if you had bought petrol.

Figure 1

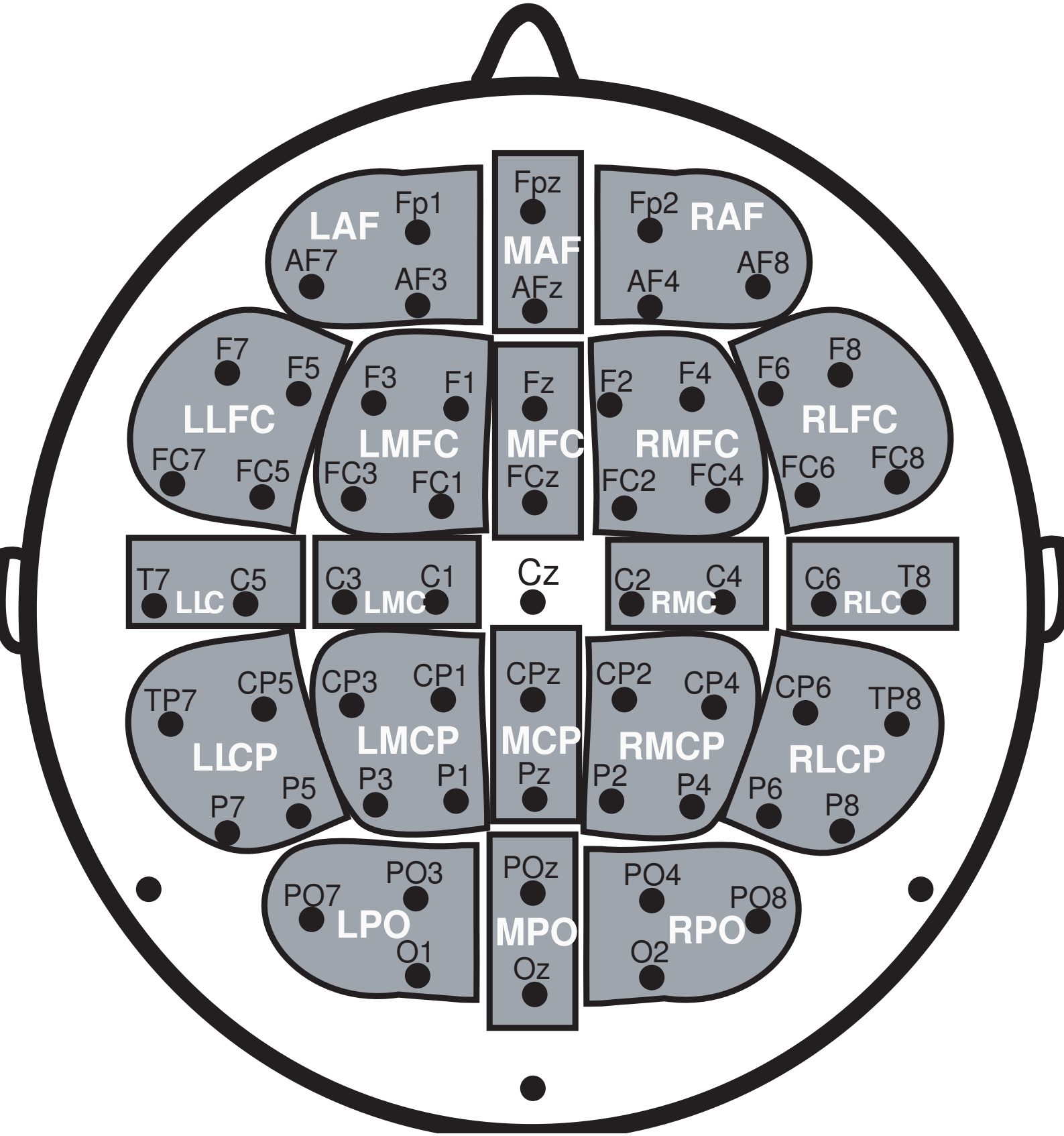
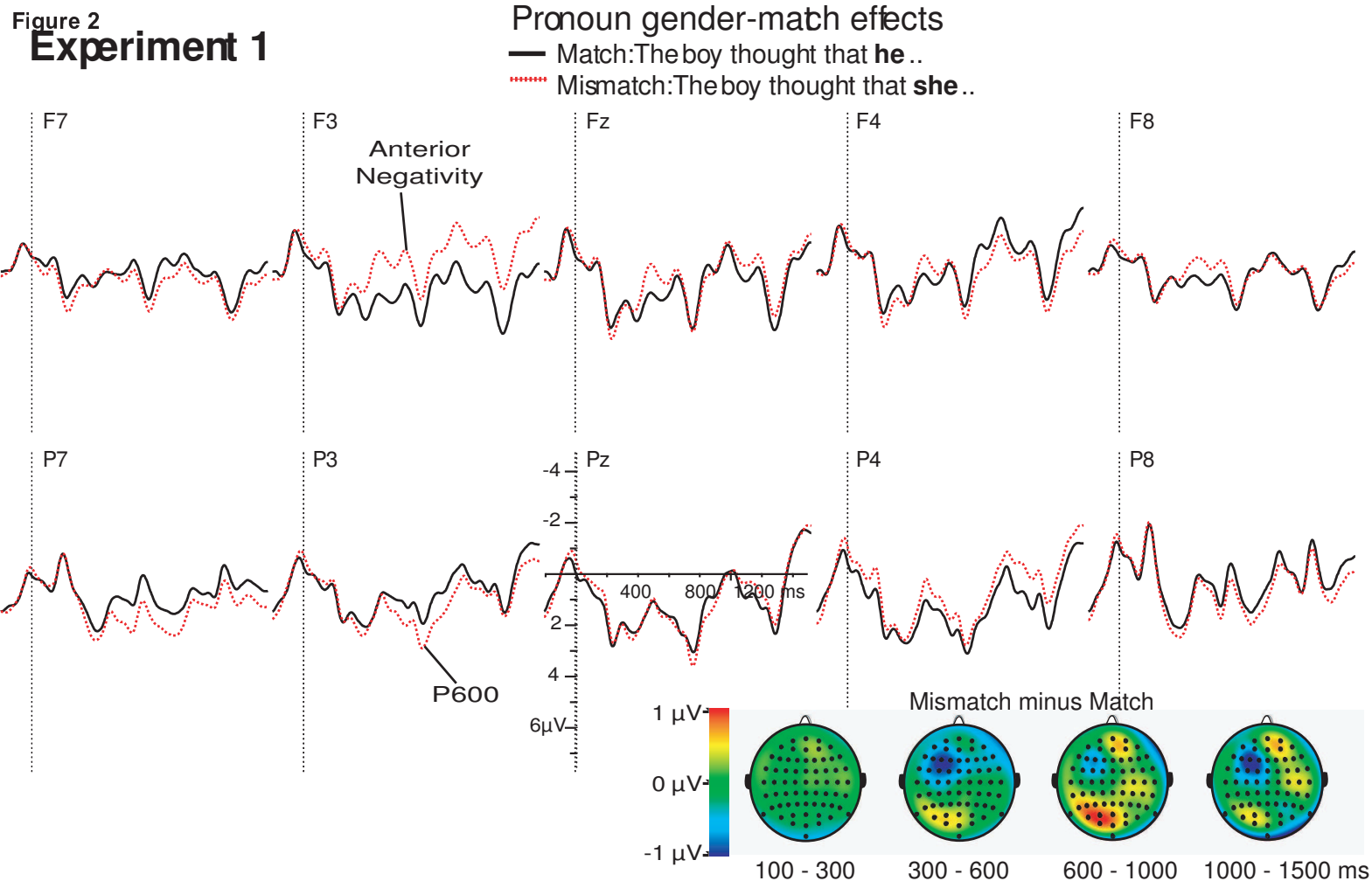
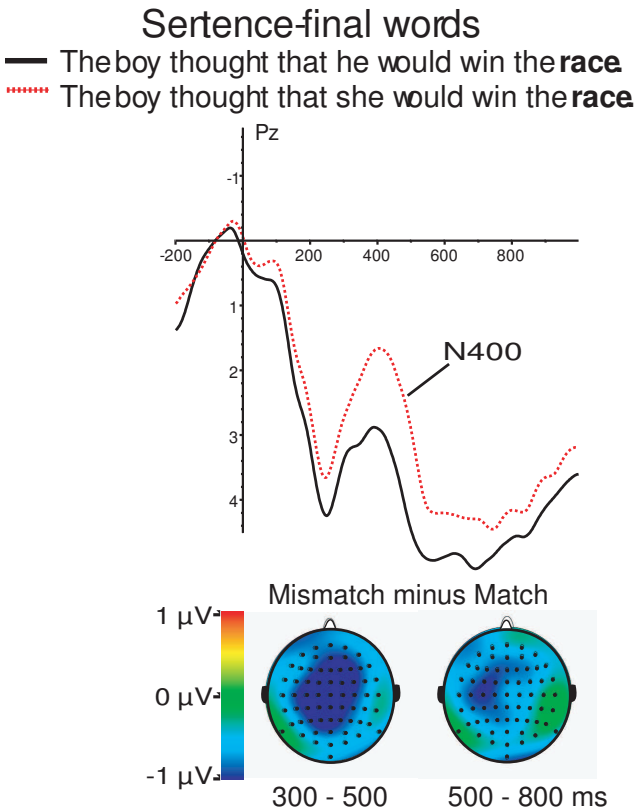

LAF
Fp1
AF7
AF3
Fpz
MAF
AFz
Fp2
RAF
AF4
AF8
F7
F5
LLFC
FC7
FC5
F3
F1
LMFC
FC3
FC1
Fz
MFC
FCz
F2
F4
RMFC
FC2
FC4
F6
F8
RLFC
FC6
FC8
T7
C5
LLC
C3
C1
LMC
Cz
C2
C4
RMC
C6
T8
RLC
TP7
CP5
LLCP
P7
P5
CP3
CP1
LMCP
P3
P1
CPz
MCP
Pz
CP2
CP4
RMCP
P2
P4
CP6
TP8
RLCP
P6
P8
PO7
PO3
LPO
O1
POz
MPO
Oz
PO4
PO8
RPO
O2

Figure 2

Experiment 1





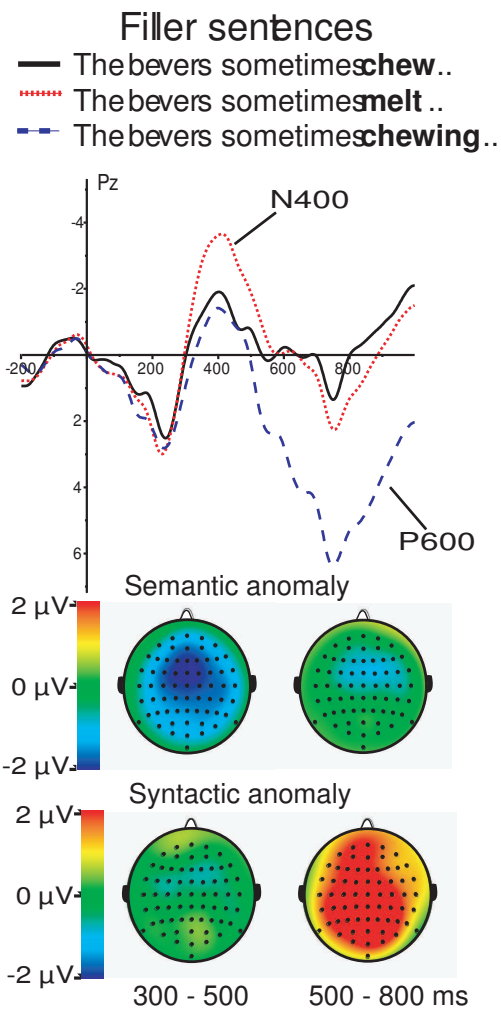


Figure 3

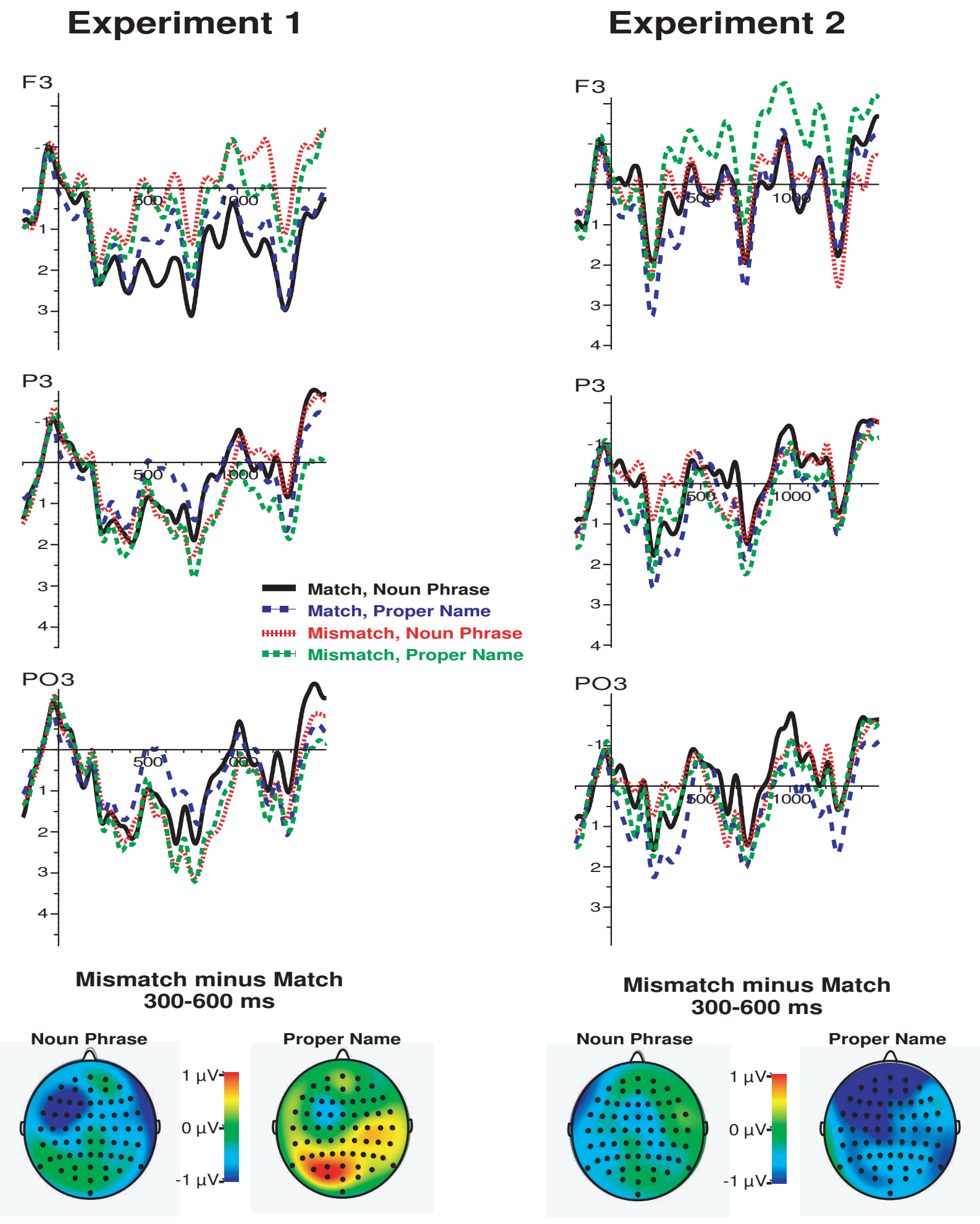

Experiment 1
Experiment 2
F3
P3
PO3
500
1000
Match, Noun Phrase
Match, Proper Name
Mismatch, Noun Phrase
Mismatch, Proper Name
Mismatch minus Match
300-600 ms
Noun Phrase
Proper Name
1 μV
0 μV
-1 μV

Figure 4

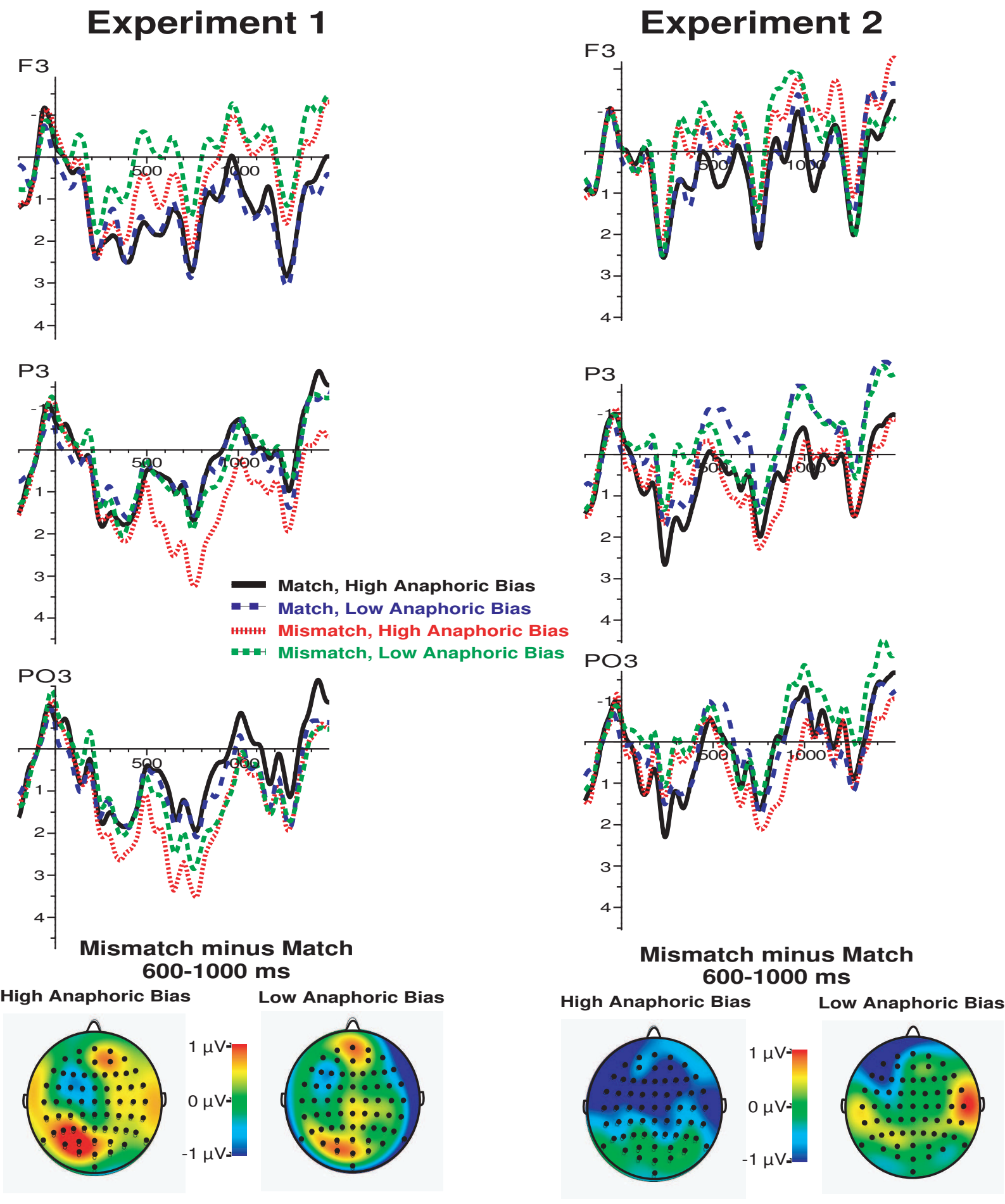
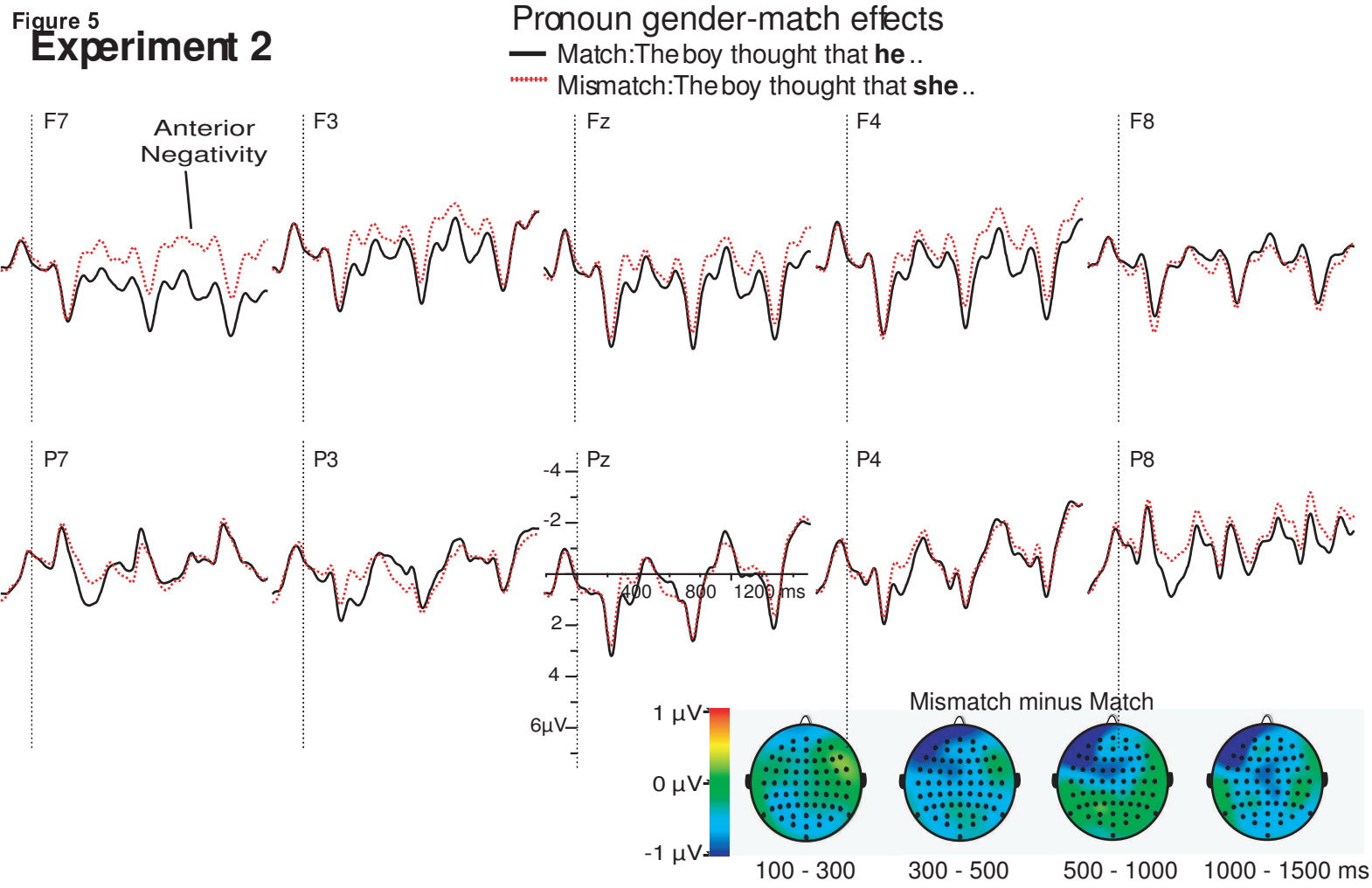
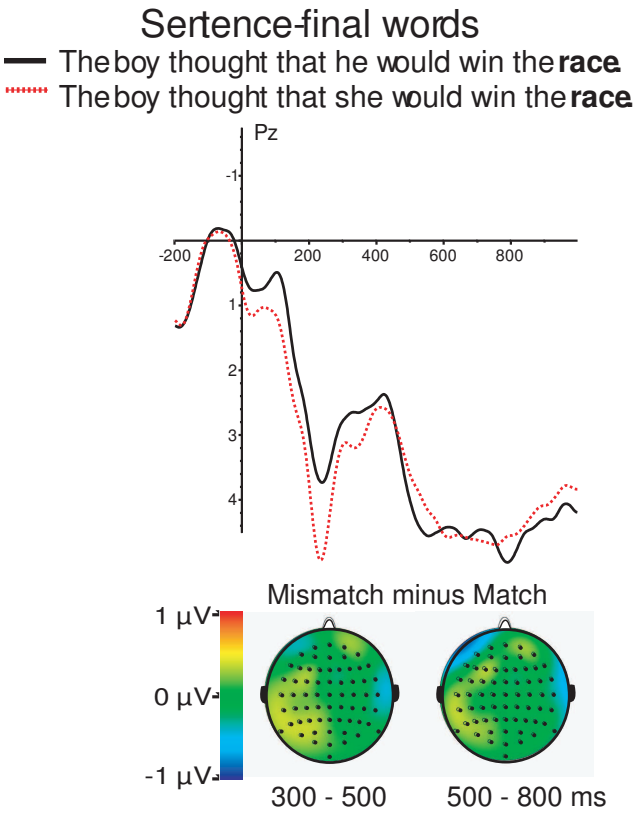

Experiment 1
Experiment 2
F3
P3
PO3
500
1000
Match, High Anaphoric Bias
Match, Low Anaphoric Bias
Mismatch, High Anaphoric Bias
Mismatch, Low Anaphoric Bias
Mismatch minus Match
600-1000 ms
High Anaphoric Bias
Low Anaphoric Bias
1 μV
0 μV
-1 μV

Figure 5

Experiment 2





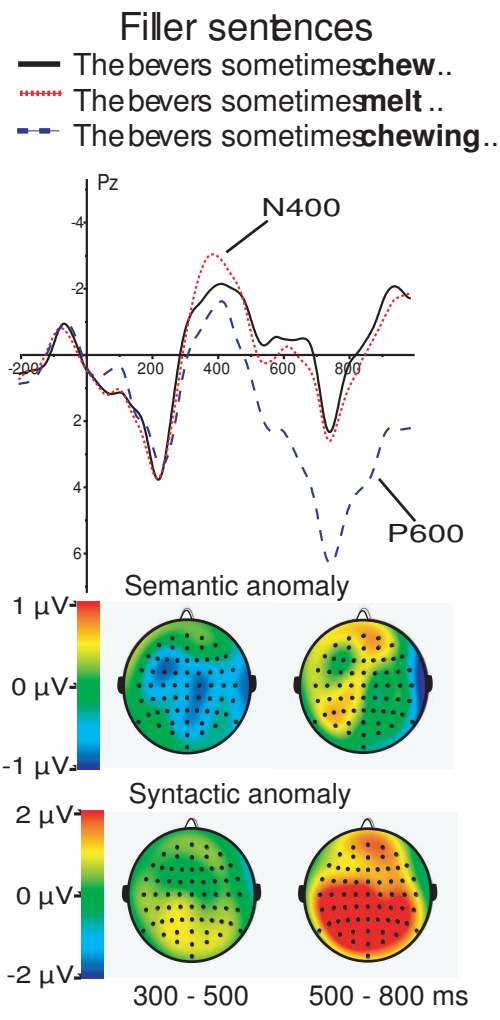
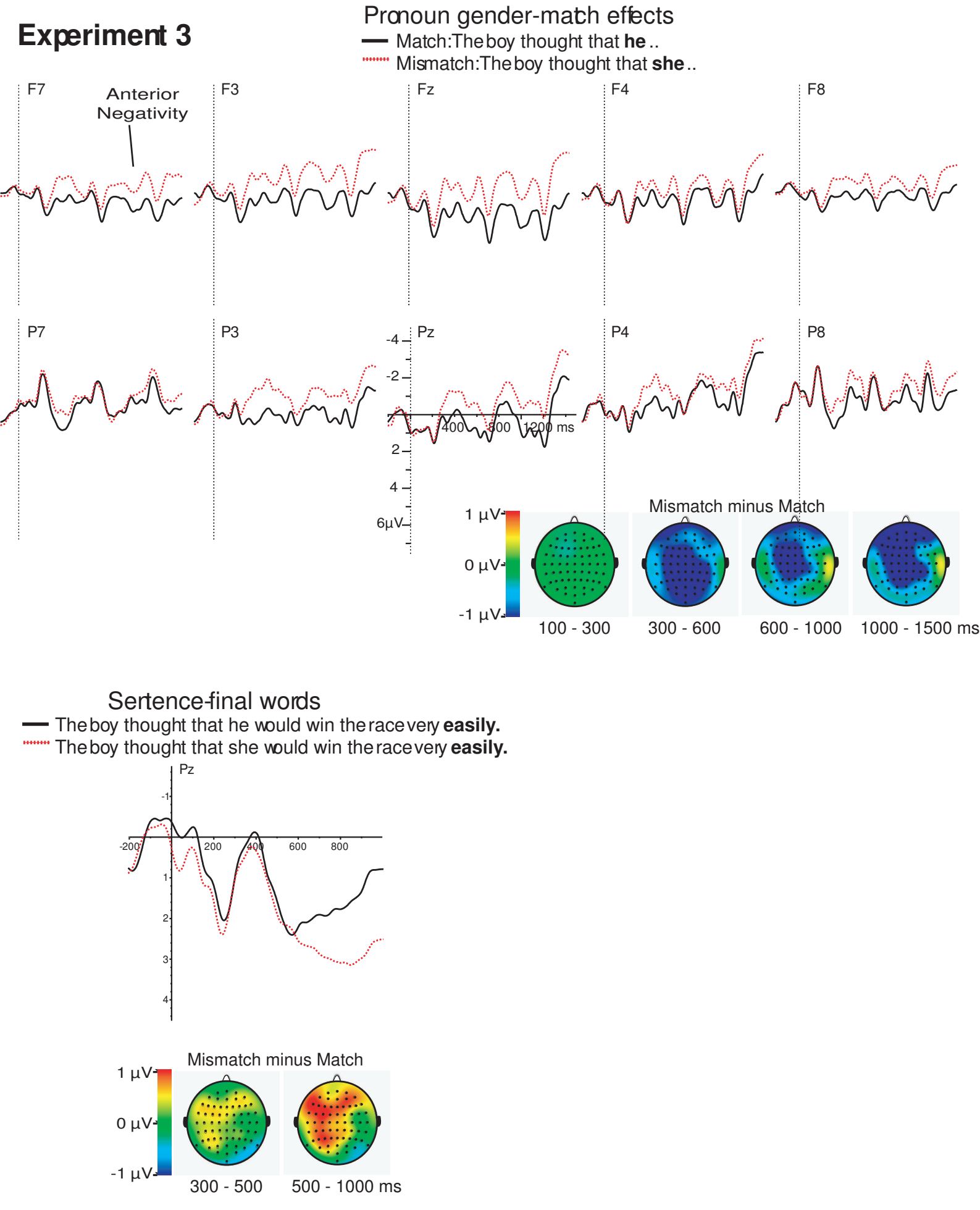


Figure 6


Experiment 3
Pronoun gender-match effects
Match: The boy thought that **he** ..
Mismatch: The boy thought that **she** ..
F7
Anterior Negativity
F3
Fz
F4
F8
P7
P3
Pz
P4
P8
-4
-2
2
4
6µV
400
800
1200 ms
Mismatch minus Match
1 µV
0 µV
-1 µV
100 - 300
300 - 600
600 - 1000
1000 - 1500 ms
Sentence-final words
The boy thought that he would win the race very **easily.**
The boy thought that she would win the race very **easily.**
Pz
-1
-200
200
400
600
800
1
2
3
4
Mismatch minus Match
1 µV
0 µV
-1 µV
300 - 500
500 - 1000 ms

This supplement contains the statistical analysis of the filler sentences in Experiment 1 and 2, followed by the analysis of the pronoun data of all experiments in a between-experiments analysis.

EXPERIMENT 1

Filler sentences

300-500 ms: A 3(Condition: control, semantic violation, syntactic violation) repeated measures ANOVA using all electrodes showed a robust effect of condition across all electrodes ($F_{2,46}$ = 10.9, p < .001): semantic violations elicited more negative ERPs compared to control sentences (M= -1.41, $S.E.$ = .36, p < .001) and syntactic violations (M= -1.2, $S.E.$ = .29, p < .001). No reliable distributional effects were observed, with the exception of a condition by hemisphere interaction in the crossline analysis ($F_{6,138}$ = 3.05, p < .05), which indicated that the effect of condition was not robust in the LLC-ROI but was the three other crossline ROIs.

500-800 ms: The three conditions elicited different ERPs across all electrodes ($F_{2,46}$ = 19.6, p < .001): syntactic violations elicited more positive ERPs than control sentences (M= 2.1 , $S.E.$ = .50, p < .001) and semantic violations (M= 2.4, $S.E.$ = .34, p < .001).

Reading Span results

There was neither a significant correlation between reading span and the N400 effect (LMCP-ROI, 300-500 ms; r =.12, *ns.*) nor one with the P600 effect (LMCP-ROI, 500-800 ms; r =.22, *ns.*)

EXPERIMENT 2

Filler sentences

300-500 ms: The effect of condition was modulated by anteriority in the medial analysis ($F_{2,36}$ = 3.4, p < .05) and marginally so in the midline analysis ($F_{6,108}$ = 2.1, p < .1). Follow-up analysis in the medial ROIs showed that the conditions differed significantly in the LMCP-ROI ($F_{2,36}$ = 3.3, p < .05): semantic violations elicited larger N400s the syntactic violations (M = -1.43, $S.E.$ = .63, p < .05). The midline follow-ups showed that difference between the three conditions strongest in the MPO-ROI ($F_{2,36}$ = 5.1, p < .05): semantic violations elicited larger N400s than control sentences (M = -.93, $S.E.$ = .40, p < .05) and syntactic violations (M = -1.49, $S.E.$ = .45, p < .005).

500-800 ms: The three conditions elicited different ERPs when using all electrodes ($F_{2,36}$ = 8.6, p = .001): syntactic violations elicited more positive ERPs than control sentences (M= 1.9 , $S.E.$ = .57, p < .005) and semantic violations (M= 1.7, $S.E.$ = .51, p < .005). Differences between the conditions depended on anteriority in the lateral ($F_{6,108}$ = 3.3, p < .05) and medial analysis ($F_{2,36}$ = 5.3, p = .01). For lateral ROIs, effects of condition were statistically significant in LLCP/RLCP-ROIs ($F_{2,36}$ = 12.8, p < .001) and LPO/RPO-ROIs ($F_{2,36}$ = 13.7, p < .001). In both these ROIs, syntactic violations elicited more positive ERPs than control sentences and semantic violations (all Ps < .005). In the medial analysis, effect of condition reached marginal significance at anterior ROIs ($F_{2,36}$ = 2.8, p < .1), but was robust at posterior ROIs ($F_{2,36}$ = 10.7, p < .001). Follow-up at posterior ROIs showed a similar effect as observed for lateral posterior ROIs: syntactic violations elicited more positive ERPs than control sentences and semantic violations (all Ps < .005).

Reading Span results

Higher reading span was associated with a larger, more negative N400 effect (LMCP-ROI, 300-500 ms; r = -.53, p = .02). No such effect was found for the P600 effect (LMCP-ROI, 500-800 ms; r =.22, *ns.*)

Combined-experiments and between-experiment analysis of the pronoun ERP data

To test for between-experiment differences, the lateral, medial, midline and crossline analyses were repeated with experiment as the between-subject factor with three levels.

300-600 ms: In the distributional analyses (lateral, medial, midline and crossline) mismatch elicited more negative ERPs across all electrodes (all *F*s > 11), but an anteriority by hemisphere by match effect was found in the medial analysis ($F_{2,75} = 4.5$, $p < .05$): mismatch elicited more negative ERPs in all quadrants (all Fs > 5) but these effects were largest at anterior channels. Match effects differed between experiments at midline ROIs only ($F_{2,72} = 4.3$, $p < .05$). The latter finding is consistent with the more widespread effect in this time window in Experiment 3, including a robust difference at midline electrodes, compared to Experiment 1 and 2.

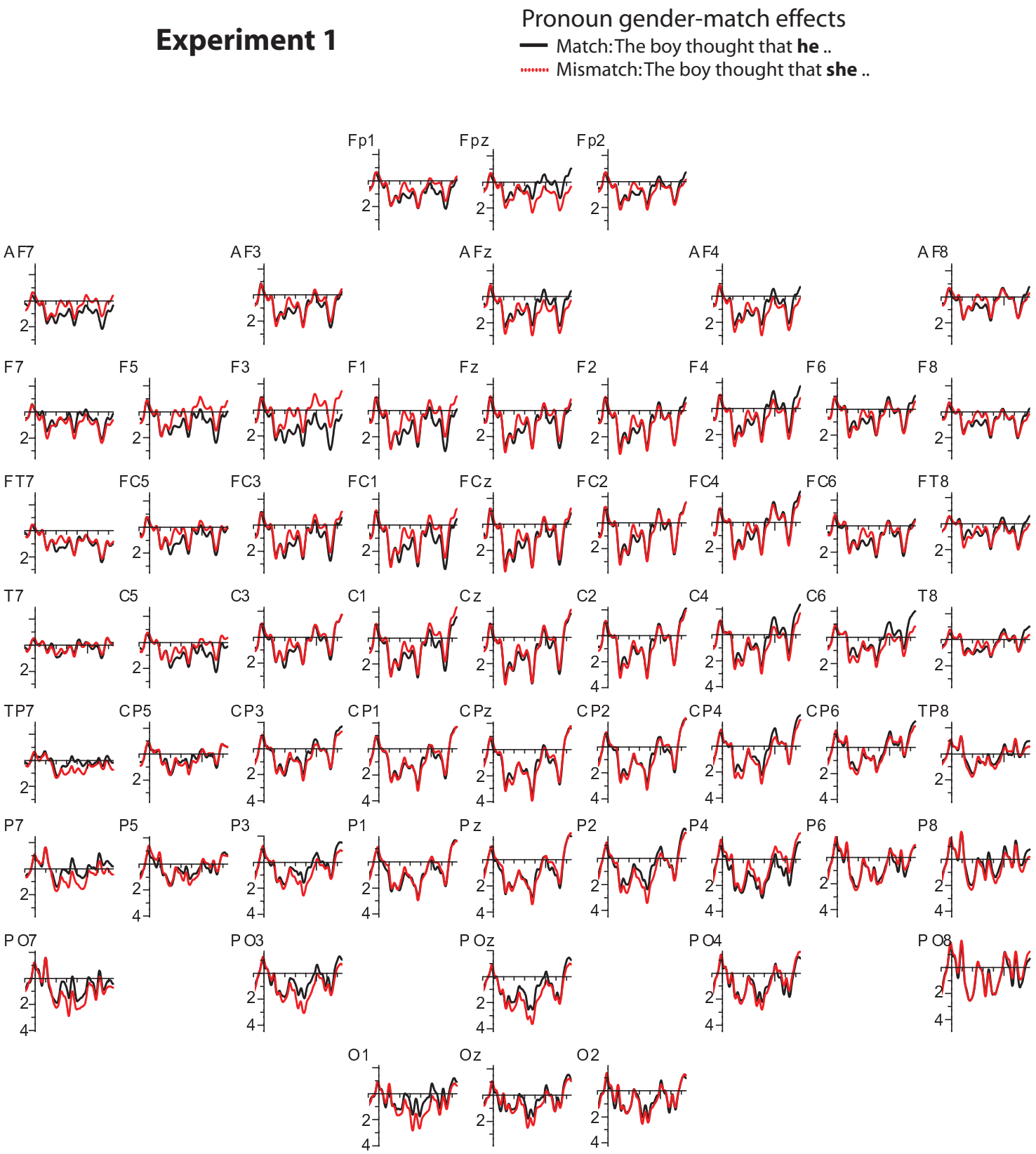
600-1000 ms: A robust match by anteriority interaction was found in the lateral analysis ($F_{2,72} = 4.3$, $p < .05$) and midline analysis ($F_{2,72} = 4.1$, $p < .05$). Mismatch elicited more negative ERPs only at the left anterior (LAF and LLFC) ROIs (both *F*s > 6). In the medial analysis, a robust match by anteriority by hemisphere interaction ($F_{2,72} = 4.7$, $p < .05$) was observed, as mismatch elicited more negative ERPs at the LMFC-ROI. A significant between-experiment effect was found in the midline analysis only ($F_{2,75} = 4.1$, $p < .05$). Similar to the previous time window, this midline differences is consistent with the Nref being strongly lateralized in Experiment 1 (possibly due to P600 component overlap) while being more widespread in Experiment 3.

1000-1500 ms: Mismatch elicited more negative ERPs across all electrodes ($F_{1,75} = 4.3$, $p < .05$), in the medial analysis ($F_{1,75} = 4.5$, $p < .05$) and in the midline analysis ($F_{1,75} = 4.8$, $p < .05$). The medial analysis also revealed match by anteriority by hemisphere

interaction effect ($F_{1,75} = 4.7$, $p < .05$): mismatch elicited more negative ERPs at the LMFC-ROI ($F_{1,75} = 9.2$, $p < .005$). No robust between-experiment effects were observed.

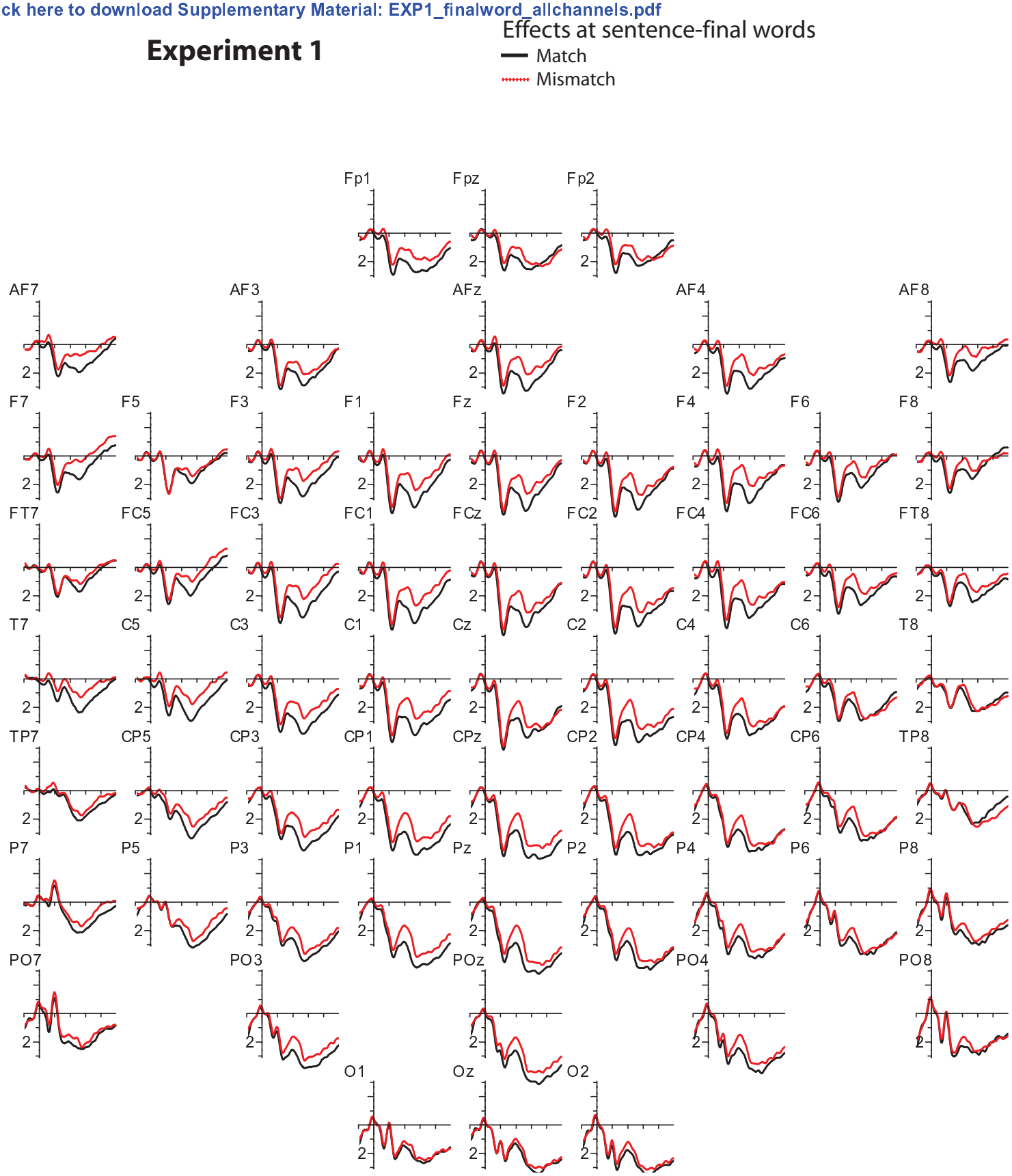
Exp1. pronoun-elicited ERPs at all channels

Click here to download Supplementary Material: EXP1_allchannels.pdf



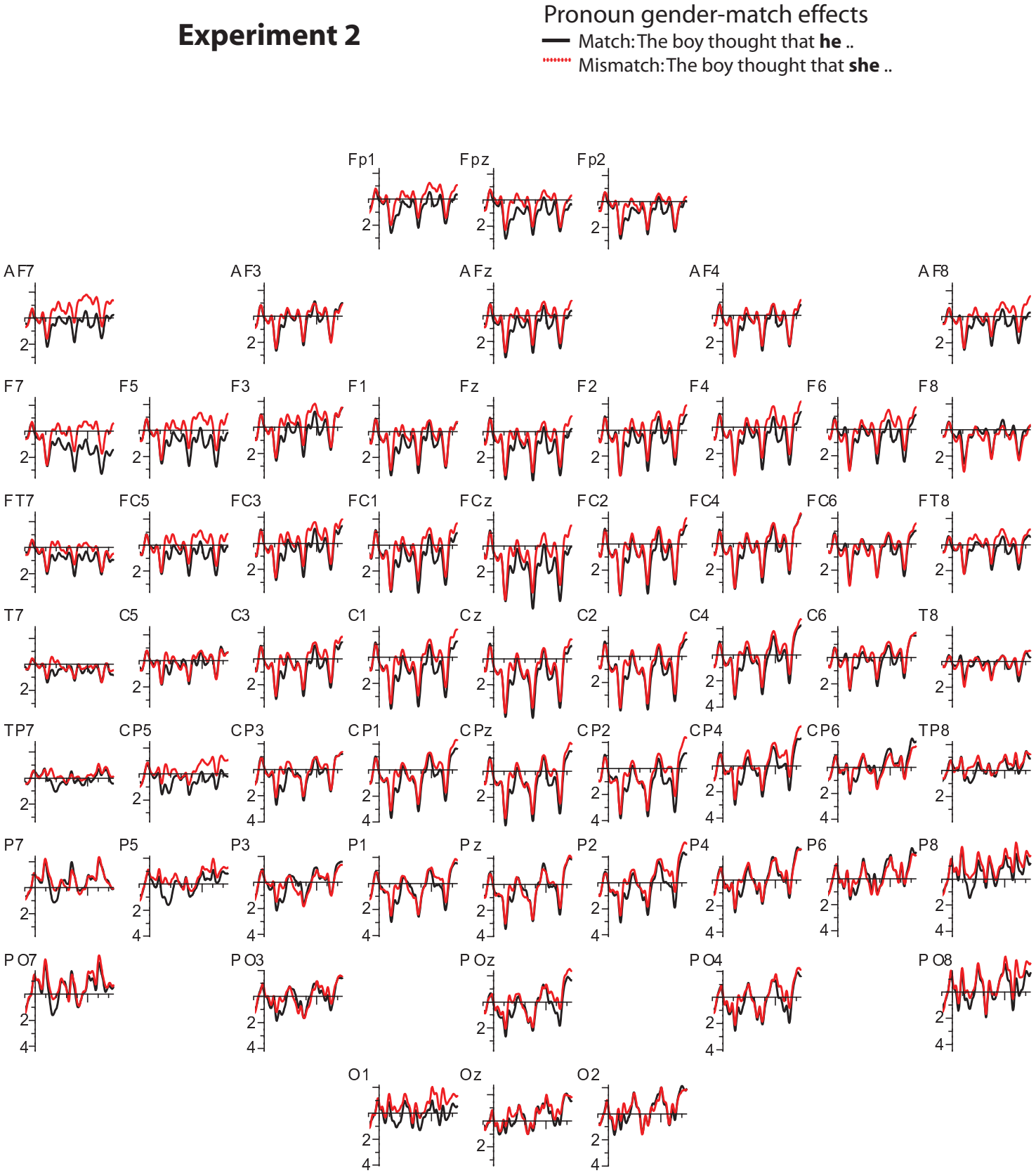
Exp1. sentencefinalword-elicited ERPs at all channels

Click here to download Supplementary Material: EXP1_finalword_allchannels.pdf



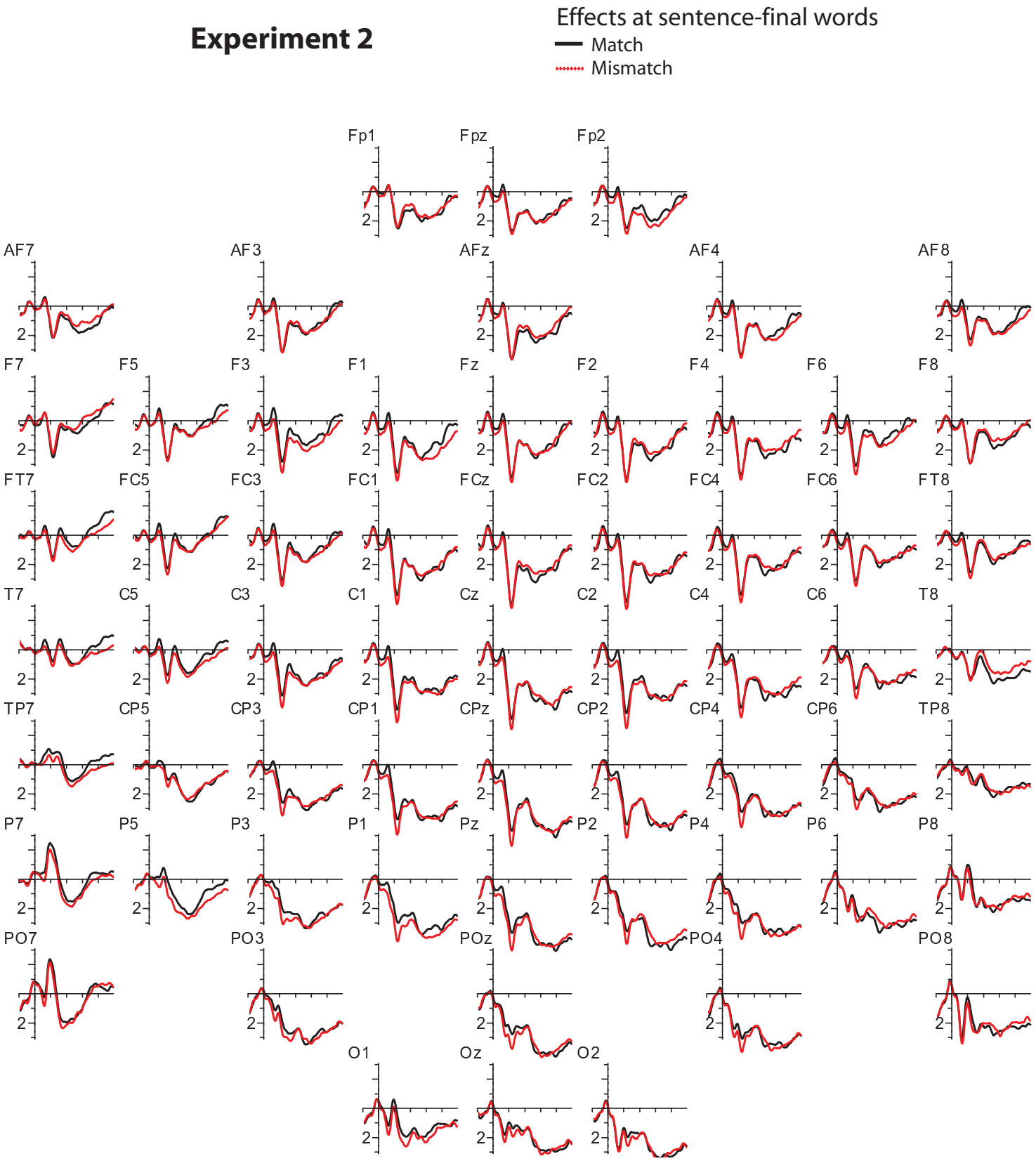
Exp2. pronoun-elicited ERPs at all channels

Click here to download Supplementary Material: EXP2_allchannels.pdf



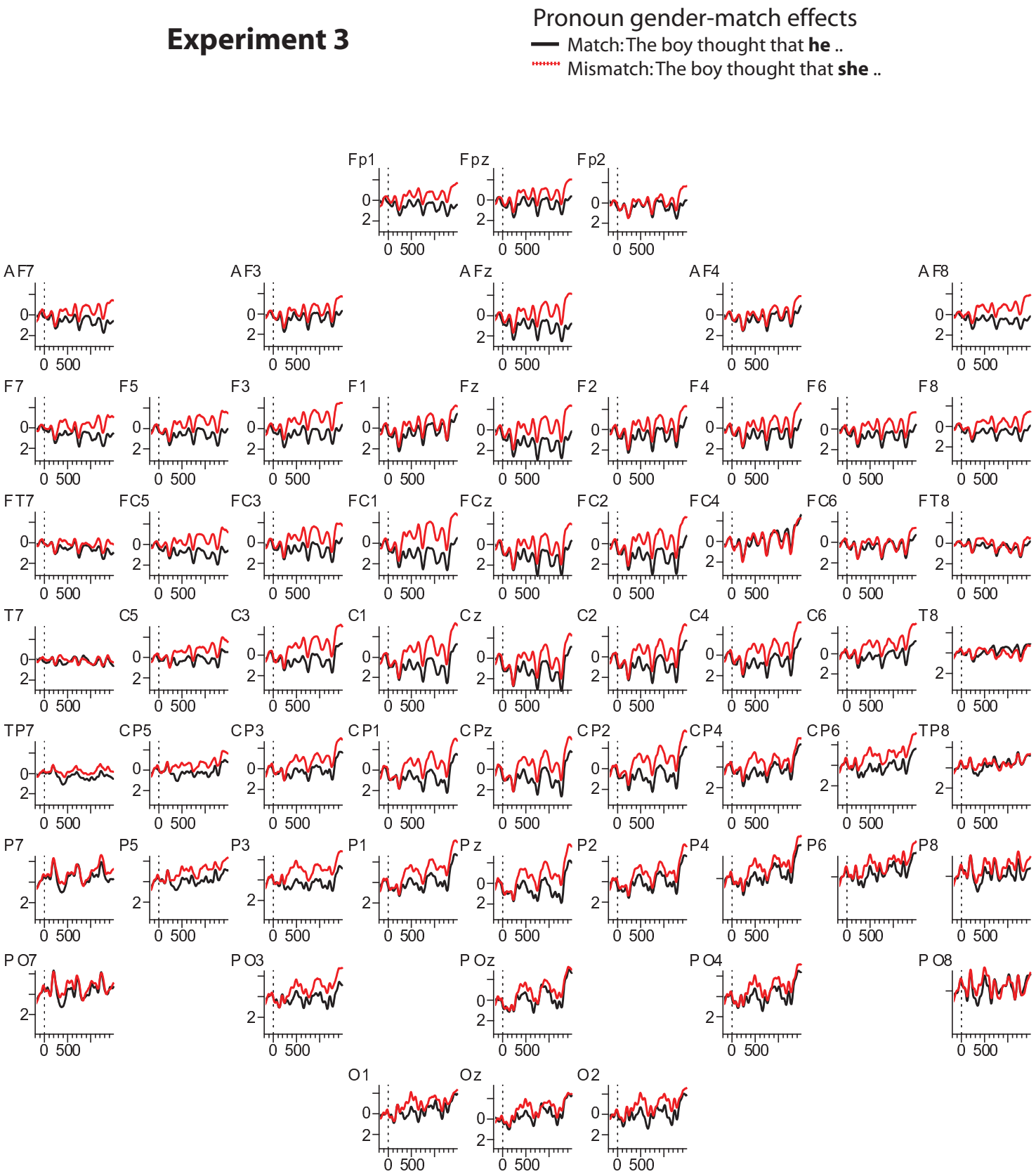
Exp2. sentencefinalword-elicited ERPs at all channels

Click here to download Supplementary Material: EXP2_finalword_allchannels.pdf



Exp3. pronoun-elicited ERPs at all channels

Click here to download Supplementary Material: EXP3_allchannels.pdf



Exp3. sentencefinalword-elicited ERPs at all channels

Click here to download Supplementary Material: EXP3_finalword_allchannels.pdf